

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

Query Match 68.5%; Score 150; DB 20; Length 221;
 Best Local Similarity 100.0%; Pred. No. 7.4e-135;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLLYRRPVNTVP 60
 DB 1 MLLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLLYRRPVNTVP 60
 QY 61 SYPGNTYDTGLPSYPMILTSRGPVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 DB 61 SYPGNTYDTGLPSYPMILTSRGPVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 QY 121 AAAPAPPIAABPAAAPLTPTVAAPPA 150
 DB 121 AAAPAPPIAABPAAAPLTPTVAAPPA 150

RESULT 7

AA30653
 ID AAM30653 standard; Protein; 219 AA.

AA30653;
 AC AAM30653;

12-APR-1999 (first entry)

Human secreted protein clone cp16 1 protein.

Human; secreted protein; nutritional activity; cytokine; vaccine;
 cell proliferation; differentiation; immune stimulation; suppression;
 haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
 chemokinetic; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
 tumour invasion suppression; tumour inhibition.

Homo sapiens.

MO9901466-A1.

14-JAN-1999.

01-JUL-1998; 98MO-US13813.

27-OCT-1997; 97US-0958304.

02-JUL-1997; 97US-0887195.

(GEMT) GENETICS INST INC.

Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racine LA;
 Spaulding V, Treacy M;

WPI; 1999-105994/09.

N-PSDB; AAB80740.

New polynucleotides encoding secreted human proteins - are derived
 from human foetal brain, adult testes, adult brain, foetal kidney,
 adult salivary gland, or adult blood cDNA libraries, useful as, e.g.

potential vaccines

Claim 24; Page 71-72; 107pp; English.

The present sequence represents a human secreted protein from clone
 cp16.1, deposited as ATCC 98482. Human secreted protein clone
 polynucleotides and proteins are predicted to have biological
 activities which would make them suitable for treating, preventing or
 ameliorating medical conditions in humans and animals. Suggested
 activities include nutritional activity, cytokine and cell
 proliferation/differentiation activity, immune stimulating (e.g. as
 vaccines) or suppressing activity, haematopoiesis regulating activity,
 tissue growth activity, activin/inhibin activity, chemotactic/
 chemokinetic activity, haemostatic and thrombolytic activity, receptor/
 ligand activity, anti-inflammatory activity, cadherin/tumour invasion
 suppressor activity, and tumour inhibition activity. The polynucleotides
 are also stated to be useful for gene therapy.

XX Sequence 219 AA;
 SQ Query Match 63.9%; Score 140; DB 20; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-125;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLLYRRPVNTVP 60
 DB 1 MLLIMACTIVCAFAKRRPPFPGEDNDGDHPLPSLNIPYGINLPPLLYRRPVNTVP 60
 QY 61 SYPGNTYDTGLPSYPMILTSRGPVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 DB 61 SYPGNTYDTGLPSYPMILTSRGPVYHIRGFLATOLNVPPLPRGPFVPSRFFSA 120
 QY 121 AAAPAPPIAABPAAAPLTPTVAAPPA 140
 DB 121 AAAPAPPIAABPAAAPLTPTVAAPPA 140

RESULT 8

ABB64202
 ID ABB64202 standard; Protein; 75 AA.

ABB64202;
 AC ABB64202;

26-MAR-2002 (first entry)

Drosophila melanogaster polypeptide SEQ ID NO 19398.

Drosophila; developmental biology; cell signalling; insecticide;
 pharmaceutical.

Drosophila melanogaster.

WO200171042-A2.

27-SEP-2001.

23-MAR-2001; 2001WO-US092331.

23-MAR-2000; 2000US-191637P.

11-JUL-2000; 2000US-061415P.

(PEKE) PE CORP NY.

Venter JC, Adams M, Li PMD, Myers EM;
 WPI; 2001-656860/75.

N-PSDB; ABL08305.

New isolated nucleic acid detection reagent for detecting 1000 or more
 genes from Drosophila and for elucidating cell signalling and cell-cell
 interactions -

Disclosure; SEQ ID/NO 19398; 21pp + Sequence Listing; English.

The invention relates to an isolated nucleic acid detection reagent
 capable of detecting 1000 or more genes from Drosophila. The invention is
 useful in developmental biology and in elucidating cell signalling and
 cell-cell interactions in higher eukaryotes for the development of
 insecticides, pharmaceuticals and pharmaceutical drugs. The invention
 discloses genetic DNA sequences (ABL16176-ABL30511), expressed DNA
 sequences (ABL01840-ABL16175) and the encoded proteins
 (ABBS7737-ABBS72072).
 The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from WIPO
 at ftp.wipo.int/pub/published_pct_sequences.

Query Match 4.1%; Score 9; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.91;

DR WPI; 2001-041424/05.
 DR N-PSDB; AAE27729.
 XX Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 PS Claim 2; Page 130; 165pp; English.
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated TPTs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 XX
 SO Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLIACIVCAVAPARRRRPPPTGDDNDGHLPSLNPYGRRLPPLYYRPNVTP 60
 DB 1 MKLLIACIVCAVAPARRRRPPPTGDDNDGHLPSLNPYGRRLPPLYYRPNVTP 60
 QY 61 SYGNTYTDGLSPYMLTSPGPPYVYHIGFPLATQLVNPPPRGPPFVPSRPFSA 120
 DB 61 SYGNTYTDGLSPYMLTSPGPPYVYHIGFPLATQLVNPPPRGPPFVPSRPFSA 120
 QY 121 AAAPAAPPIAEPAAAAPLTATPVAAEPAAAGAVAAEPAAEAVGAEPAEAPVAAEPAA 180
 DB 121 AAAPAAPPIAEPAAAAPLTATPVAAEPAAAGAVAAEPAAEAVGAEPAEAPVAAEPAA 180
 QY 181 EAPVGEPAEAPSPAPATAPKAPAPPPSPLEQANQ 219
 DB 181 EAPVGEPAEAPSPAPATAPKAPAPPPSPLEQANQ 219
 RESULT 2
 AAE27862
 ID AAE27862 standard; Protein; 219 AA.
 XX
 AC AAE27862;
 XX
 DT 27-DEC-2002 (first entry)
 XX
 DE Human zsig63 protein.
 XX
 KM Human; secreted salivary protein; zsig63 protein; host defense protein;
 KM immune modulating factor; antipathogenic; cell-cell signaling molecule;
 KM growth factor; cytokine; growth factor hormone activity; dental caries;
 KM infection; tooth decay; periodontal disease; gastrointestinal disease;
 KM thrush; urinary tract infection; vaginal infection; diabetes; obesity;
 KM anti-inflammatory; chronic tissue damage; lung dysfunction; restenosis;
 KM gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KM forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT 1..15
 FT Peptide /label= Signal_peptide
 FT 16..219
 FT Protein /note= "Mature human zsig63 protein"
 FT 14..21
 FT Region /note= "Hydrophilic region"
 FT 16..37
 FT Domain /note= "Domain 1"
 FT 17..33
 FT Region /note= "Antigenic epitope"
 FT 24..30
 FT Region /note= "Hydrophilic region"
 FT 38..126
 FT Domain

FT /note= "Domain 2"
 FT 65..73
 FT Region /note= "Antigenic epitope"
 FT 103..108
 FT Region /note= "Antigenic epitope"
 FT 124..133
 FT Region /note= "Repeat 1"
 FT 127..219
 FT Domain /note= "Domain 3"
 FT 134..138
 FT Region /note= "Repeat 2"
 FT 139..143
 FT Region /note= "Repeat 3"
 FT 144..148
 FT Region /note= "Repeat 4"
 FT 149..153
 FT Region /note= "Repeat 5"
 FT 154..158
 FT Region /note= "Repeat 6"
 FT 159..163
 FT Region /note= "Repeat 7"
 FT 164..168
 FT Region /note= "Repeat 8"
 FT 169..173
 FT Region /note= "Repeat 9"
 FT 174..178
 FT Region /note= "Repeat 10"
 FT 179..183
 FT Region /note= "Repeat 11"
 FT 184..188
 FT Region /note= "Repeat 12"
 FT 187..192
 FT Region /note= "Hydrophilic region"
 FT 189..193
 FT Region /note= "Repeat 13"
 FT 190..197
 FT Region /note= "Antigenic epitope"
 FT 194..198
 FT Region /note= "Repeat 14"
 FT 199..203
 FT Region /note= "Repeat 15"
 FT 202..215
 FT Region /note= "Antigenic epitope"
 FT 204..208
 FT Region /note= "Repeat 16"
 XX
 FN US2002090677-A1.
 XX
 PD 11-JUL-2002.
 XX
 PP 03-AUG-2001; 2001US-0923236.
 XX
 PR 17-MAR-1999; 99US-124820P.
 PR 17-MAR-2000; 2000US-0527345.
 XX
 PA (ADLER/) ADLER D A.
 PA (SHEP/) SHEPPARD P O.
 XX
 PI Adler DA, Sheppard PO;
 XX
 DR WPI; 2002-642378/69.
 DR N-PSDB; AAD45050.
 XX
 PT Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 PT agent for treating microbial infection, dental caries, periodontal
 PT disease, thrush gastrointestinal disease, and for aiding digestion -
 XX
 PS Claim 10; Page 28; 33pp; English.
 CC The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agents and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

CC antipathogenic polypeptide, cell-cell signalling molecule, growth factor,
 CC cytokine, or as secreted extracellular matrix associated proteins with
 CC growth factor hormone activity. It is useful for treating conditions
 CC associated with pathological microbes, including bacterial, fungal and
 CC viral infections, for treating dental caries (tooth decay), periodontal
 CC disease, thrush and gastrointestinal diseases, for treating urinary tract
 CC infection, vaginal infection and for preventing infection in skin and
 CC other epithelial wounds. zsig63 is useful for establishing normal
 CC microflora and protect against pathogenic colonisation and invasion, for
 CC treating chronic tissue damage e.g. damage in extremities associated
 CC with diabetes and useful as anti-inflammatory agents. It is useful as a
 CC marker of lung dysfunction, salivary gland dysfunction, or dysfunction of
 CC prostate gland. It is also therapeutically useful for aiding digestion.
 CC Polynucleotides of the invention are used in gene therapy for increasing
 CC or inhibiting zsig63 activity, for detecting abnormalities on human
 CC chromosome 4 associated with disease or other human traits and as
 CC diagnostics in forensic DNA profiling. Sequences of the invention are
 CC useful for stimulating proliferation or differentiation of cardiac
 CC myocytes, for proliferation or differentiation of adipocytes and for
 CC inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.
 CC The present sequence is human zsig63 protein. zsig63 gene is located
 CC on chromosome 4.

XX Sequence 219 AA;

Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1,1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTICVAPARKRRPFPGEDDNDGHPHPSINIRYGINLPPLYYRPNVTP 60
 DB 1 MKLLIMACTICVAPARKRRPFPGEDDNDGHPHPSINIRYGINLPPLYYRPNVTP 60
 QY 61 SYPGNTYDTGSPFWILTSRPFYVYHNGFPLATOLNPLPRGPFVPPSPRPSA 120
 DB 61 SYPGNTYDTGSPFWILTSRPFYVYHNGFPLATOLNPLPRGPFVPPSPRPSA 120
 QY 121 AAAPAAPPIAEPAAAPLTAIPVAAEPDAGAPVAAEPAAEPVGAEPAAEPVAAEPDA 180
 DB 121 AAAPAAPPIAEPAAAPLTAIPVAAEPDAGAPVAAEPAAEPVGAEPAAEPVAAEPDA 180
 QY 181 EAPVGVPEAAEPSPAPAPATKAPAPBPSPSPLEQANQ 219
 DB 181 EAPVGVPEAAEPSPAPAPATKAPAPBPSPSPLEQANQ 219

RESULT 3

ABG31608 ID ABG31608 standard; Protein; 219 AA.

XX ABG31608;

DT 15-NOV-2002 (first entry)

DE Human secreted salivary protein zsig63.

KW Human; secreted salivary protein; zsig63; immunogen; zsig63-cytokine;
 KW antibody-cytokine; in vivo killing; pathological microbe; bacteria;
 KW fungal; viral; infection; salivary gland; anti-microbial; dental caries;
 KW tooth decay; periodontal disease; thrush; gastrointestinal disease;
 KW urinary tract infection; vaginal infection; skin infection; microflora;
 KW epithelial wound; pathogenic colonisation; invasion; pro-inflammatory;
 KW chronic tissue damage; vascular system; diabetes; anti-inflammatory;
 KW incompetent immune system; AIDS; acquired immunodeficiency syndrome;
 KW chemotherapy; radiation treatment; lung infection; cystic fibrosis;
 KW digestion; chromosome 4.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Region 1..15 /note= "Fusion protein peptide, specifically claimed in claim 18"

FT Region 1..219 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 9..204 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 14..19 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 16..21 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 16..37 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 17..33 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 24..29 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 24..33 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 25..30 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 38..126 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 66..73 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 103..108 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 109..215 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 124..133 /label= Repeat_1
 FT 127..219 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT Region 134..138 /label= Repeat_2
 FT 139..143 /label= Repeat_3
 FT 144..148 /label= Repeat_4
 FT 149..153 /label= Repeat_5
 FT 154..158 /label= Repeat_6
 FT 159..163 /label= Repeat_7
 FT 164..168 /label= Repeat_8
 FT 169..173 /label= Repeat_9
 FT 174..178 /label= Repeat_10
 FT 179..183 /label= Repeat_11
 FT 184..188 /label= Repeat_12
 FT 187..192 /note= "Antigenic peptide, specifically claimed in claim 15"
 FT 189..193

FT Region /label= Repeat_13
 FT 190..197 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 194..198
 FT /label= Repeat_14
 FT Region 199..203
 FT /label= Repeat_15
 FT 202..215 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 204..208
 FT /label= Repeat_16
 FT US2002081701-A1.
 XX
 XX 27-JUN-2002.
 XX
 XX 03-AUG-2001; 2001US-0922480.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX 17-MAR-2000; 2000US-0527345.
 XX
 XX (ADLER/) ADLER D A.
 XX (SHEP/) SHEPPARD P O.
 XX
 XX Adler DA, Sheppard PO;
 XX
 XX MPI: 2002-635468/68.
 XX N-PSDB; A852633, A852634.
 XX
 XX Novel secreted salivary protein, zsig63 and polynucleotide encoding it
 XX useful for treating microbial infections, inflammatory conditions, -
 XX dental caries and lung infections associated with cystic fibrosis
 XX
 XX Claim 10; Page 28; 33pp; English.
 XX
 XX The present invention relates to a new secreted salivary protein, zsig63.
 XX The invention is useful for detecting in a test sample, the presence of
 XX an antagonist or agonist of zsig63 protein activity. The invention is
 XX also useful as an immunogen for producing an antibody to zsig63
 XX polypeptide. zsig63-cytokine fusion proteins or antibody-cytokine fusion
 XX protein are useful for enhancing in vivo killing of target tissues.
 XX Pharmaceutical composition comprising purified zsig63 polypeptide are
 XX useful in the treatment of conditions associated with pathological
 XX microbes, including bacterial, fungal and viral infections. High
 XX expression of zsig63 in salivary gland suggests that anti-microbial
 XX polypeptides are useful for treatment of dental caries (tooth decay),
 XX periodontal disease, thrush and gastrointestinal disease. Other
 XX applications can be used in urinary tract infections, vaginal infections,
 XX prevention of infection in skin and other epithelial wounds. The
 XX polypeptides can be used to establish normal microflora and protect
 XX against pathogenic colonisation and invasion. The invention is useful
 XX when pro-inflammatory activity is desired. Applications for
 XX such pro-inflammatory activity include the treatment of chronic tissue
 XX damage, particularly in areas having a limited or damaged vascular system
 XX e.g., damage in extremities associated with diabetes. Antagonists to
 XX zsig63 polypeptides may be useful as anti-inflammatory agents. The
 XX invention is useful for the treatment of patients having incompetent
 XX immune system, such as AIDS (acquired immunodeficiency syndrome) patients
 XX or individuals that have undergone chemotherapy, radiation treatment. The
 XX invention is also useful for the treatment of lung infections associated
 XX with cystic fibrosis and its agonists or antagonists are useful for
 XX aiding digestion. The present amino acid sequence represents the human
 XX secreted salivary protein zsig63 of the invention. This sequence is
 XX encoded by the human zsig63 gene located on chromosome 4.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTVCVAFARKRRFPETIGDDDDGGHPLPSINTPYGINLLPPLLYRRPVNTVP 60
 Db 1 MKLLIMACTVCVAFARKRRFPETIGDDDDGGHPLPSINTPYGINLLPPLLYRRPVNTVP 60
 QY 61 SYPGNTTYDTGCLPSFPMILTSRPFYVYHIRSFPLATQLVNPPPLPRGRPFVPSRFSFA 120
 Db 61 SYPGNTTYDTGCLPSFPMILTSRPFYVYHIRSFPLATQLVNPPPLPRGRPFVPSRFSFA 120
 QY 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 Db 121 AAAPAAPPIAAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAA 180
 QY 181 BAPGVBPAAEPSPAPAPATAPAAEPBPSPSLBOAQ 219
 Db 181 BAPGVBPAAEPSPAPAPATAPAAEPBPSPSLBOAQ 219
 RESULT 4
 AAU74536
 ID AAU74536 standard; Protein; 219 AA.
 XX
 XX AAU74536;
 XX
 XX 23-APR-2002 (first entry)
 XX
 XX Human zsig63 polypeptide.
 XX
 XX Human; zsig63; chromosome 4q12-q413; salivary protein; antimicrobial;
 XX microbial infection; tooth decay; periodontal disease; thrush; emphysema;
 XX gastrointestinal disease; urinary tract infection; vaginal infection;
 XX skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
 XX acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
 XX chronic bronchitis; gene therapy; protein therapy.
 XX
 OS Homo sapiens.
 XX
 XX US6331413-B1.
 XX
 XX 18-DEC-2001.
 XX
 XX 17-MAR-2000; 2000US-0527345.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX
 XX (ZYMO) ZYMOGENETICS INC.
 XX
 XX Adler DA, Sheppard PO;
 XX
 XX MPI: 2002-096707/13.
 XX N-PSDB; A8520591.
 XX
 PT Polynucleotides encoding salivary proteins useful as anti-microbial
 XX agents -
 XX
 XX Claim 1; Column 49-52; 29pp; English.
 XX
 XX The invention relates to a polynucleotide derived from the 4q12-4q13
 XX region of human chromosome 4 and encoding a zsig63 polypeptide, a
 XX secreted salivary protein with anti-microbial activity. Due to their
 XX microbial activity, the sequences can be used in the study of microbial
 XX infections, e.g. for recombinant production of anti-microbial proteins.
 XX The sequences can be used in the treatment of tooth decay, periodontal
 XX disease, thrush, gastrointestinal disease, urinary tract infections,
 XX vaginal infections, skin infections, epithelial wounds, chronic tissue
 XX damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
 XX infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
 XX represents human zsig63.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

DR N-PSDB; ABX93594.
 XX Novel isolated zsig63 polypeptide, member of the adhesin family, useful
 PT for treating dental caries, periodontal disease, thrush,
 PT gastrointestinal disease, urinary tract infections, vaginal infections,
 PT skin infections
 XX
 XX Claim 10; Page 27-28; 32pp; English.
 XX
 CC The invention relates to an isolated zsig63 polypeptide comprising at
 CC least 90% identity to an amino acid sequence which comprises domain 1 of
 CC zsig63, domain 2, domain 3, mature zsig63 and full length zsig63. Also
 CC included are the polynucleotide encoding zsig63, a zsig63 expression
 CC vector, a cultured cell comprising the vector and expressing the protein,
 CC a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37,
 CC 38-126, 127-219 or 16-219 of zsig63 and an additional protein), using a
 CC zsig63 reporter gene construct to identify zsig63 agonists, and
 CC producing an anti-zsig63 antibody using zsig63 immunogenic peptides,
 CC zsig63 is useful for detecting in a test sample, the presence of
 CC antagonist of zsig63 protein activity. Zsig63 has antimicrobial
 CC activity and since exhibits high expression in salivary gland, can be
 CC used for treating dental caries, periodontal disease, thrush, and
 CC gastrointestinal disease, urinary tract infections, vaginal infections,
 CC skin infections and other epithelial wounds. The polypeptides can be
 CC used to establish normal microflora and protect against pathogenic
 CC colonization and invasion. Zsig63 can also be used for providing
 CC pro-inflammatory activity for treating chronic, tissue damage
 CC particularly in areas having limited or damaged vascular system, e.g.
 CC in diabetes, and for treating immunocompromised AIDS patients or in
 CC individuals that have undergone chemotherapy, radiation treatment, for
 CC treating lung infections e.g. in cystic fibrosis. Detection of zsig63
 CC polypeptide at relatively high levels in the trachea may indicate that
 CC such polypeptides may serve as a marker of lung dysfunction. Zsig63 is
 CC also useful in diagnosing conditions associated with salivary gland or
 CC lung dysfunction including salivary gland carcinoma, Pneumocystis carinii
 CC infection, emphysema, chronic bronchitis, prostate dysfunction such
 CC as prostate adenocarcinoma, aiding digestion, and as components of
 CC defined cell culture media and may be used to replace serum that is
 CC commonly used in culture. The DNA is useful in gene therapy applications
 CC to increase or inhibit zsig63 activity, and for detecting abnormalities
 CC on human chromosome 4 (e.g. 4q12-4q13), associated with dentingeneis
 CC imperfecta, and dentin dysplasia type II). Zsig63 is an adhesin family
 CC member. The present sequence represents human zsig63.
 CC
 XX
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 219; DB 24; Length 219;
 Best Local Similarity 100.0%; Pred. No. 1.1e-200;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MKLLWACTVCAFAARRRPPFGEDNDGHPHLSLNPYGIKRLPPLYYRPVNTVP 60
 DB 1 MKLLWACTVCAFAARRRPPFGEDNDGHPHLSLNPYGIKRLPPLYYRPVNTVP 60
 QY 61 SYGNTYTTDGLSPYWLITSPGPPYVYHIRGPLATQLVNPLPRGPPVPPSPFSA 120
 DB 61 SYGNTYTTDGLSPYWLITSPGPPYVYHIRGPLATQLVNPLPRGPPVPPSPFSA 120
 QY 121 AAAPAPPTAAAPAAAPLATATVAAPAAAGAVAAEPAAEAVGAEPAAPAAEPAA 180
 DB 121 AAAPAPPTAAAPAAAPLATATVAAPAAAGAVAAEPAAEAVGAEPAAPAAEPAA 180
 QY 181 EAPVGEPAAPAEPSPEPATAPKAPAEPPSPLEQANQ 219
 DB 181 EAPVGEPAAPAEPSPEPATAPKAPAEPPSPLEQANQ 219
 RESULT 6
 ID AAY19472 standard; Protein; 221 AA.
 XX
 AC AAY19472;
 XX

DT 14-JUL-1999 (first entry)
 XX
 DB Amino acid sequence of a human secreted protein.
 XX
 KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
 KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
 KW immune system disease; autoimmune disease; hepatic disease;
 KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
 KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
 KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
 KW lung disease; thymus disease; digestive disorder; endocrine disorder;
 KW infection; AIDS.
 XX
 OS Homo sapiens.
 XX
 PN MO9922243-AL.
 XX
 PD 06-MAY-1999.
 XX
 PP 23-OCT-1998; 98WO-US22376.
 XX
 PR 24-OCT-1997; 97US-0063387.
 PR 24-OCT-1997; 97US-0062784.
 PR 24-OCT-1997; 97US-0063088.
 PR 24-OCT-1997; 97US-0063090.
 PR 24-OCT-1997; 97US-0063091.
 PR 24-OCT-1997; 97US-0063092.
 PR 24-OCT-1997; 97US-0063097.
 PR 24-OCT-1997; 97US-0063098.
 PR 24-OCT-1997; 97US-0063100.
 PR 24-OCT-1997; 97US-0063101.
 PR 24-OCT-1997; 97US-0063109.
 PR 24-OCT-1997; 97US-0063110.
 PR 24-OCT-1997; 97US-0063111.
 PR 24-OCT-1997; 97US-0063148.
 PR 24-OCT-1997; 97US-0063386.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Carter KC, Duan DR, Ebner R, Entress GA,
 PI Feng P, Florence C, Florence KA, Greene JM, Janat F,
 PI Kay H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA,
 PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
 XX
 DR WPI; 1999-303069/25.
 DR N-PSDB; AAX61352.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PS Claim 11; Page 401-402; 546pp; English.
 XX
 CC The specification describes cDNA sequences (AAX61322-X61470) encoding
 CC human secreted proteins (AAY19442-Y19590). The polynucleotides and their
 CC corresponding secreted polypeptides are useful for preventing, treating
 CC or ameliorating medical conditions, e.g. by protein or gene therapy.
 CC Pathological conditions can also be diagnosed by determining the amount
 CC of the polypeptides in a sample or by determining the presence of
 CC mutations in the polynucleotides. Specific uses are described for each
 CC of the polynucleotides, based on which tissues they are most highly
 CC expressed in, and include developing products for the diagnosis or
 CC treatment of cancer, tumours, neurodegenerative disorders, developmental
 CC abnormalities and fetal deficiencies, blood disorders, leukemias,
 CC diseases of the immune system, autoimmune diseases, hepatic and renal
 CC disease, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
 CC and cognitive disorders, schizophrenia, prostate diseases, obesity,
 CC disorders involving osteoclasts such as osteoporosis, arthritis or
 CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
 CC disorders, infections and AIDS. The polypeptides are also useful for
 CC identifying their binding partners.
 XX
 SQ Sequence 221 AA;
 XX

Query Match 68.5%; Score 150; DB 20; Length 221;
 Best Local Similarity 100.0%; Pred. No. 7.4e-135;
 Matches 150; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACTVCVAFARRRPPFIGEDNDGDHPLHPSLNTIPYGRNLPPPLYRPNVTVP 60
 DB 1 MKLLMACTVCVAFARRRPPFIGEDNDGDHPLHPSLNTIPYGRNLPPPLYRPNVTVP 60
 QY 61 SYPGNTYDTGLPSYPMILTSRGPYYVYHIRGFLATQLVNPPPLPRGFPVPSRPFSA 120
 DB 61 SYPGNTYDTGLPSYPMILTSRGPYYVYHIRGFLATQLVNPPPLPRGFPVPSRPFSA 120
 QY 121 AAAPAPPIAAEPAAAPLTPVAEPAA 150
 DB 121 AAAPAPPIAAEPAAAPLTPVAEPAA 150

RESULT 7
 AAW30653
 ID AAW30653 standard; Protein; 219 AA.
 AC AAW30653;
 DT 12-APR-1999 (first entry)
 DE Human secreted protein clone cp16 1 protein.
 KW Human; secreted protein; nutritional activity; cytokine; vaccine;
 KW cell proliferation; differentiation; immune stimulation; suppression;
 KW haemopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
 KW chemokine; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
 KW tumour invasion suppression; tumour inhibition.
 OS Homo sapiens.
 XX MO9901466-A1.
 XX 14-JAN-1999.
 PD
 PF 01-JUL-1998; 98MO-US13813.
 PR 27-OCT-1997; 97US-0958304.
 PR 02-JUL-1997; 97US-0887195.
 XX (GENY) GENETICS INST INC.
 PA
 PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Racie LA;
 PI Spaulding V, Treacy M;
 DR WPI; 1999-105994/09.
 DR N-PSDB; AAV80740.
 XX
 PT New polynucleotides encoding secreted human proteins - are derived
 PT from human foetal brain, adult testes, adult brain, foetal kidney,
 PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.,
 PT potential vaccines
 PS
 XX Claim 24; Page 71-72; 107pp; English.
 CC The present sequence represents a human secreted protein from clone
 CC cp16 1, deposited as ATCC 98482. Human secreted protein clone
 CC polynucleotides and proteins are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals. Suggested
 CC activities include nutritional activity, cytokine and cell
 CC proliferation/differentiation activity, immune stimulating (e.g. as
 CC vaccines) or suppressing activity, haematopoiesis regulating activity,
 CC tissue growth activity, activin/inhibin activity, chemotactic/
 CC chemokine activity, haemostatic and thrombolytic activity, receptor/
 CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
 CC suppressor activity, and tumour inhibition activity. The polynucleotides
 CC are also stated to be useful for gene therapy.

XX SQ Sequence 219 AA;
 Query Match 63.9%; Score 140; DB 20; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-125;
 Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLMACTVCVAFARRRPPFIGEDNDGDHPLHPSLNTIPYGRNLPPPLYRPNVTVP 60
 DB 1 MKLLMACTVCVAFARRRPPFIGEDNDGDHPLHPSLNTIPYGRNLPPPLYRPNVTVP 60
 QY 61 SYPGNTYDTGLPSYPMILTSRGPYYVYHIRGFLATQLVNPPPLPRGFPVPSRPFSA 120
 DB 61 SYPGNTYDTGLPSYPMILTSRGPYYVYHIRGFLATQLVNPPPLPRGFPVPSRPFSA 120
 QY 121 AAAPAPPIAAEPAAAPLTPVAEPAA 140
 DB 121 AAAPAPPIAAEPAAAPLTPVAEPAA 140

RESULT 8
 ABB64202
 ID ABB64202 standard; Protein; 75 AA.
 AC ABB64202;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 19398.
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 KW Drosophila melanogaster.
 OS Drosophila melanogaster.
 XX MO200171042-A2.
 XX 27-SEP-2001.
 PD
 PF 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA
 PI Venter JC, Adams M, Li PMD, Myers EW;
 PI WPI; 2001-656860/75.
 DR N-PSDB; ABL08305.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS
 XX Disclosure; SEQ ID NO 19398; 21pp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL10511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 75 AA;
 Query Match 4.1%; Score 9; DB 22; Length 75;
 Best Local Similarity 100.0%; Pred. No. 0.91;

2730933

```

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 120 AAAAPAAP 128
    |||||
Db 45 AAAAPAAP 53

RESULT 9
AAW03688
ID AAW03688 standard; Protein; 322 AA.
XX
AC AAW03688;
XX
DT 09-MAR-1997 (first entry)
XX
DE Leishmania chagasi acidic ribosomal antigen LcP0.
XX
KW Leishmania chagasi; acidic ribosomal antigen; LcP0;
XX epitope; K39.
XX
OS Leishmania chagasi.
XX
FH Key Location/Qualifiers
FT CDS 30..1202
FT /*tag= a
XX
PN MO963414-A2.
XX
PD 24-OCT-1996.
XX
PF 19-APR-1996; 96MO-US05472.
XX
PR 21-APR-1995; 95US-0428414.
XX
PA (COR1-) CORIXA CORP.
XX
PI Reed SG;
XX
PI 1996-485884/48.
XX
DR N-PSDB; AAT42164.
XX
PT New Leishmania acidic ribosomal P-protein family polypeptide - used
PT to develop prods. for diagnosis, detection and protection against
PT Leishmania infections
XX
PS Disclosure; Page 29-32; 76pp; English.
XX
CC Compounds including polypeptides that contain at least an epitope of
CC the L. chagasi acidic ribosomal antigen LcP0 are useful in a variety
CC of immunoassays for detecting Leishmania infection. Portions of
CC LcP0 (AAT42164) contg. at least the 17 C-terminal amino acids (AAT42165)
CC have been found to generate a signal in an ELISA that is equivalent
CC to that generated by the full length LcP0. A combination
CC polypeptide may also be used, comprising an LcP0 epitope along with
CC an epitope derived from the Leishmania K39 antigen (AAT42166), pref.
CC the K39 repeat unit antigen having the sequence given in AAW03690.
XX
SQ Sequence 322 AA;

Query Match 4.1%; Score 9; DB 17; Length 322;
Best Local Similarity 100.0%; Pred. No. 3.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAP 138
    |||||
Db 289 AAEPAAAP 297

RESULT 10
ABB65430
ID ABB65430 standard; Protein; 728 AA.
XX
AC ABB65430;

```

```

XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 23082.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
XX
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PERK ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR N-PSDB; ABL09533.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 23082; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (AAB57737-AB872072).
XX
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 728 AA;

Query Match 4.1%; Score 9; DB 22; Length 728;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAAP 127
    |||||
Db 418 SAAAPAAP 426

RESULT 11
AAB76532
ID AAB76532 standard; Protein; 1874 AA.
XX
AC AAB76532;
XX
DT 11-APR-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein SEQ ID NO:46.
XX
KW Corynebacterium glutamicum; brevbacterium lacticofementum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering.
XX
OS Corynebacterium glutamicum.
XX

```

PN WO200100805-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00926.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031454.
 PR 08-JUL-1999; 99DE-1031478.
 PR 08-JUL-1999; 99DE-1031563.
 PR 09-JUL-1999; 99DE-1032122.
 PR 09-JUL-1999; 99DE-1032124.
 PR 09-JUL-1999; 99DE-1032125.
 PR 09-JUL-1999; 99DE-1032128.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032182.
 PR 09-JUL-1999; 99DE-1032190.
 PR 09-JUL-1999; 99DE-1032191.
 PR 09-JUL-1999; 99DE-1032209.
 PR 09-JUL-1999; 99DE-1032212.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032228.
 PR 09-JUL-1999; 99DE-1032229.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032927.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1033006.
 PR 27-AUG-1999; 99DE-1040764.
 PR 27-AUG-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040766.
 PR 27-AUG-1999; 99DE-1040830.
 PR 27-AUG-1999; 99DE-1040831.
 PR 27-AUG-1999; 99DE-1040832.
 PR 27-AUG-1999; 99DE-1040833.
 PR 31-AUG-1999; 99DE-1041378.
 PR 31-AUG-1999; 99DE-1041379.
 PR 31-AUG-1999; 99DE-1041385.
 PR 03-SEP-1999; 99DE-1042077.
 PR 03-SEP-1999; 99DE-1042078.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042088.
 XX
 PA (BADT) BASF AG.
 XX
 PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
 XX
 DR WPI: 2001-071486/08.
 DR N-PSDB; AAF67765.
 XX
 PT Corynebacterium glutamicum nucleic acids encoding membrane construction
 PT and membrane transport proteins or their portions, useful for typing or
 PT identifying C. glutamicum or related bacteria, and as markers for
 PT transformation -
 XX
 XX Claim 20; Page 202-207; 119pp; English.
 XX
 XX AAF67743 to AAF68080 encode the Corynebacterium glutamicum membrane
 CC construction and membrane transport (MCT) proteins given in AAB76510 to
 CC AAB76847. The MCT nucleic acids and proteins are useful in the
 CC identification of microorganisms which can be used to produce fine
 CC chemicals, for modulating fine chemical production in C. glutamicum or
 CC related bacteria (e.g. Brevibacterium lactofermentum), the typing or
 CC identification of C. glutamicum or related bacteria, as reference points
 CC for mapping C. glutamicum genome, and as markers for transformation.
 CC AAF68082 and AAF68082 represent sequencing primers which are used in an
 CC example from the present invention.
 CC
 XX
 SQ Sequence 1874 AA;
 XX
 Query Match 4.1%; Score 9; DB 22; Length 1874;
 Best Local Similarity 100.0%; Pred. No. 13;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAEAPVAA 176
 DB 568 PAAEAPVAA 576
 RESULT 12
 AAG90680 standard; Protein; 2969 AA.
 ID AAG90680
 AC AAG90680;
 XX
 DT 26-SEP-2001 (first entry)
 DE C glutamicum protein fragment SEQ ID NO: 4434.
 XX
 XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KM organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP108790-A2.
 XX
 PD 20-JUN-2001.
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 XX 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 XX (KYOW) KYOWA HAKKO KOGYO KK.
 XX
 PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 XX
 XX WPI: 2001-376931/40.
 DR N-PSDB; AAF65899.
 XX
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene
 PT -
 XX
 PS Claim 17; SEQ ID NO: 4434; 246pp + Sequence listing; English.
 XX
 XX The present invention provides a number of nucleotide and protein
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of coryneform bacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Coryneform bacterium, and identifying a homologue of a gene derived
 CC from coryneform bacterium. Coryneform bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.
 CC
 XX
 SQ Sequence 2969 AA;
 XX
 Query Match 4.1%; Score 9; DB 22; Length 2969;
 Best Local Similarity 100.0%; Pred. No. 20;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 168 PAAEAPVAA 176
 DB 1663 PAAEAPVAA 1671
 RESULT 13
 ABP96450 standard; peptide; 9 AA.
 ID ABP96450

XX ABP96450; (first entry)
AC 27-MAY-2003
XX HLA class I molecule heavy chain optimal peptide #1.
DE Human leukocyte antigen; HLA; major histocompatibility complex; MHC;
XX T cell.
KM Homo sapiens.
OS Synthetic.
XX WO2003016905-A2.
PN 27-FEB-2003.
XX 15-AUG-2002; 2002WO-GE03753.
PF 16-AUG-2001; 2001GB-0020042.
PR (AVID-) AVIDEX LTD.
XX Jakobson BK;
PI WPI; 2003-268366/26.
DR Determining whether a T-cell reacts with a predetermined Major
XX Histocompatibility Complex (MHC) type, comprises contacting a sample of
PT T-cell with MHC type molecules and determining whether the MHC
PT molecules activate the T-cell -
XX Disclosure; Page 17; 56pp; English.
XX The present invention describes a method for determining whether a T-cell
CC reacts with a predetermined major histocompatibility complex (MHC) type,
CC which comprises bringing a sample comprising the T-cell into contact with
CC several molecules of the MHC type, each MHC molecule being complexed with
CC a peptide antigen whose contribution to a T-cell receptor binding the
CC MHC-peptide antigen complex is minimised, and determining whether the MHC
CC molecules cause activation of the T-cell. Also described: (1) a cloning
CC vector encoding an MHC subunit (preferably human) into which a nucleotide
CC encoding a peptide antigen sequence of interest can be inserted such that
CC expression of the vector produces a fusion protein comprising the MHC
CC subunit with the peptide antigen fused into it via a linker sequence;
CC (2) a cell transformed with the above vector; (3) a multivalent class I
CC or II MHC/peptide complex, preferably a multimer, in which the peptide
CC antigen is not recognised in a specific manner by T-cells; and (4) a kit
CC for carrying out the above method, comprising the plurality of MHC
CC molecules cited above. The method is useful in determining whether a
CC T-cell reacts with a predetermined major histocompatibility complex type,
CC and in assessing allo-specific T-cell activity, or in matching transplant
CC and donor patients, as well as in monitoring alloreactive responses
CC following a transplant operation. The present sequence represents a human
CC leukocyte antigen (HLA) class I molecule heavy chain optimal peptide,
CC which is given in the exemplification of the present invention.
XX
SQ Sequence 9 AA;
Query Match 3.7%; Score 8; DB 24; Length 9;
Best Local Similarity 100.0%; Pred. No. 9.3e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAA 137
DB 1 AAEPAAAA 8

RESULT 14
ABP80080
ID ABP80080 standard; Protein; 59 AA.
XX
AC ABP80080;

XX 07-MAR-2003 (first entry)
DT N. gonorrhoeae amino acid sequence SEQ ID 6690.
XX
DE Antibacterial; infection; vaccine; gene therapy.
XX
XX Neisseria gonorrhoeae.
OS
XX WO200279243-A2.
PN 10-OCT-2002.
XX
PD 12-FEB-2002; 2002WO-IB02069.
PF 12-FEB-2001; 2001GB-0003424.
PR (CHIR-) CHIRON SPA.
XX Fontana MR, Piazza M, Masignani V, Monaci E;
XX WPI; 2003-058415/05.
DR N-PSDB; AB241050.
XX
XX New protein from Neisseria gonorrhoeae, useful for the manufacture of a
PT medicament for treating or preventing N. gonorrhoeae infection -
XX Disclosure; Page 674; 815pp; English.
XX The present invention relates to proteins from Neisseria gonorrhoeae.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing N. gonorrhoeae
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.
XX
SQ Sequence 59 AA;
Query Match 3.7%; Score 8; DB 24; Length 59;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 PATAKPAA 205
DB 25 PATAKPAA 32

RESULT 15
AAG27598
ID AAG27598 standard; Protein; 65 AA.
XX
AC AAG27598;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 32499.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
XX
PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139464.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143547.
PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147250.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147943.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 3.7%; Score 8; DB 21; Length 65;
Best Local Similarity 100.0%; Pred. No. 7.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAADA 137
|||
Db 19 AAEPAADA 26

Search completed: January 15, 2004, 15:28:31
Job time : 42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:23:11 ; Search time 18 Seconds

(without alignments)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Sequence: 1 MCLLMACIVCAVAFARKRF.....TAKPAPEHPSPSLQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 127863 seqs, 47026705 residues

Word size : 0

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.1	322	RLA0_LEICH	P39096 leishmania
2	9	4.1	323	RLA0_LEICH	P39097 leishmania
3	9	4.1	581	HEMI_STRCO	Q6wx15 streptomyces
4	9	4.1	864	NIA_VOLCA	P36841 volvox cart
5	8	3.7	107	RLA1_LEIPE	Q66313 leishmania
6	8	3.7	110	RLA3_SCHPO	P17477 schistosoma
7	8	3.7	128	RL7_CHLBN	Q32361 chlamydia p
8	8	3.7	353	CUP5_GALMB	Q24998 galliera me
9	8	3.7	396	P53_MESAV	Q00366 mesocricetu
10	8	3.7	574	ATP2_CHIRB	P38482 chlamydomon
11	8	3.7	575	MSPA_TREMA	Q92413 treponema m
12	8	3.7	576	BCR_HELVY	Q18473 helicobas v
13	8	3.7	1477	KELC_DROME	Q04652 drosophila
14	8	3.7	2470	NTC2_MOUSE	Q35516 mus musculu
15	7	3.2	82	ANPA_PSBAM	P04002 pseudopleur
16	7	3.2	101	RL12_MERTH	P05394 methanobact
17	7	3.2	101	RL12_MERTH	Q52706 methanococ
18	7	3.2	102	RL12_MERTH	P54048 methanococ
19	7	3.2	105	RLA2_DICDI	P22683 dictyostella
20	7	3.2	107	RLA1_CHLRE	P29763 chlamydomon
21	7	3.2	109	RLA1_TRYCR	P26643 trypanosoma
22	7	3.2	110	RLA2_CRYSTO	Q61463 cryptochito
23	7	3.2	111	RA2A_MAIZE	P46252 zea mays (m
24	7	3.2	111	RL12_ABRPE	Q93949 aeropyrum p
25	7	3.2	111	RLA1_CABEL	P19193 caenorhabdi
26	7	3.2	111	RLA2_ASPFU	Q90046 aspergillus
27	7	3.2	112	RLA1_DROME	P08570 drosophila
28	7	3.2	114	RL12_HALNI	P05768 halobacteri
29	7	3.2	114	RLA1_RAT	P19944 rattus norv
30	7	3.2	114	RLA2_BIMTE	Q96799 eimeria ten
31	7	3.2	120	CU19_ARADI	P90515 araneus dia
32	7	3.2	125	PSAE_SPTOL	P12354 epinacia ol
33	7	3.2	129	DYL4_CHLRE	Q93591 chlamydomon

34	7	3.2	136	1	RL28_HUMAN	P46779 homo sapien
35	7	3.2	136	1	RL28_MOUSE	P41105 mus musculu
36	7	3.2	136	1	RL14_HUMAN	P37108 homo sapien
37	7	3.2	156	1	BCC2_PSEAR	P37799 pseudomonas
38	7	3.2	163	1	ATPD_CABEL	Q09544 caenorhabdi
39	7	3.2	168	1	ATPD_HUMAN	P30049 homo sapien
40	7	3.2	199	1	CYCY_RHOCA	Q05389 rhodobacter
41	7	3.2	204	1	RS3A_METWA	Q89052 methanosarc
42	7	3.2	208	1	RS6_MYCCE	P47336 mycoplasma
43	7	3.2	215	1	ERPD_WHEAT	P29546 triticum ae
44	7	3.2	215	1	RS6_MYCPN	P75543 mycoplasma
45	7	3.2	220	1	COAT_CMVSI	Q00467 cymbidium m

ALIGNMENTS

```

RESULT 1
RLA0_LEICH          STANDARD; PRT; 322 AA.
ID P39096;
AC P39096;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN LCPO.
OS Leishmania chagasi.
OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=44271;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISOLATE MHOW/BR/82/BA-2;
RX MEDLINE=94222525; Pubmed=7513304;
RA Skelly Y.A.W., Benson D.R., Elvasilla M., Badaro R., Burns J.M. Jr.,
RA Reed S.G.;
RT "Antigens shared by leishmania species and Trypanosoma cruzi:
RT Immunological comparison of the acidic ribosomal P0 proteins.";
RL Infect. Immun. 62:1643-1651(1994).
CC -!- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF R. COLI PROTEIN L10.
CC -!- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
CC DIMERS OF P1 AND P2.
CC -!- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: L29300; AAA29262.1; -
DR InterPro: IPR001813; 60S_Ribosomal.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00428; 60S_Ribosomal; 1.
DR Pfam: PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 322 AA; 34594 MW; 2389P125356D26D2 CRC64;

Query Match          4.1%; Score 9; DB 1; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAP 138
   |||||
Db 289 AAEPAAP 297

RESULT 2
RLA0_LEICH          STANDARD; PRT; 323 AA.
ID P39097;
AC P39097;

```

DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE 60S acidic ribosomal protein P0.
 GN L1PO-A AND L1PO-B.
 OS Leishmania infantum.
 OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
 NCBI_TaxID=5671;
 RX MEDLINE=94088674; PubMed=8264730;
 RA Soto M., Requena J.M., Alonso C.;
 RT "Isolation, characterization and analysis of the expression of the
 RL Leishmania ribosomal P0 protein genes";
 M1. Biochem. Parasitol. 61:265-274(1993).
 CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
 CC OF E. COLI PROTEIN L10.
 CC -1- SUBUNIT: P0 FORMS A PENTAMERIC COMPLEX BY INTERACTION WITH
 CC DIMERS OF P1 AND P2.
 CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X72714; CAA51264.1; -;
 DR EMBL; X72714; CAA51264.1; -;
 DR InterPro; IPR001813; 60S ribosomal.
 DR InterPro; IPR001790; Ribosomal L10.
 DR Pfam; PF00428; 60S_rribosomal; 1.
 DR Pfam; PF00466; Ribosomal_L10; 1.
 KW Ribosomal protein; Phosphorylation.
 SQ SEQUENCE 323 AA; 34771 MW; 14967BD7A439D69E CRC64;
 Query Match 4.1%; Score 9; DB 1; Length 323;
 Best Local Similarity 100.0%; Pred. No. 0.66; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 130 AAEPAAP 138
 Db 290 AAEPAAP 298
 RESULT 3
 HEMI_STRCO STANDARD; PRT; 581 AA.
 ID HEMI_STRCO STANDARD; PRT; 581 AA.
 AC Q9WX15;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Glutamyl-tRNA reductase (EC 1.2.1.-) (GluTR).
 GN HEMA OR SC03319 OR SCE68.17C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomycetes.
 NCBI_TaxID=1902;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleier H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornby T., Howarth S.,
 RA Huang C.-H., Kleier T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitsch B., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorek A., Woodward J., Barrall B.G., Parkhill J.,

RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 CC -1- CATALYTIC ACTIVITY: Glutamyl-tRNA(Glu) + NADPH = glutamate-1-
 CC semialdehyde + NADP(+) + tRNA(Glu).
 CC -1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; first step.
 CC -1- SIMILARITY: Belongs to the glutamyl-tRNA reductase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AL939116; CAB45353.1; -;
 DR FIR; T36267; T36267.
 DR HAMAP; MF_00087; atypical; 1.
 DR InterPro; IPR000343; GluTR.
 DR InterPro; IPR000594; Thif domain.
 DR Pfam; PF00745; GluTR dimer; 1.
 DR Pfam; PF05201; GluTR_N; 1.
 DR Pfam; PF05200; GluTR_NAD bind; 1.
 DR TIGRPFam; TIGR01035; hemaA; 1.
 DR PROSITE; PS00747; GluTR; 1.
 KW Porphyrin biosynthesis; Oxidoreductase; NADP; Complete proteome.
 FT DOMAIN 292 416 INSERT.
 FT ACT SITE 50 50 NUCLEOPHILE (BY SIMILARITY).
 FT ACT SITE 99 99 BASE (BY SIMILARITY).
 SQ SEQUENCE 581 AA; 60562 MW; D4B256B105AFA037 CRC64;
 Query Match 4.1%; Score 9; DB 1; Length 581;
 Best Local Similarity 100.0%; Pred. No. 1.1; Indels 0; Gaps 0;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 154 VAAEPAAEA 162
 Db 36 VAAEPAAEA 44
 RESULT 4
 NIA_VOLCA STANDARD; PRT; 864 AA.
 ID NIA_VOLCA STANDARD; PRT; 864 AA.
 AC P36841;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nitrate reductase (NADH) (EC 1.7.1.1) (NR).
 GN NITR.
 OS Nitrospira.
 OC Bacteria; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
 OC Volvocaceae; Volvox.
 NCBI_TaxID=3067;
 RX MEDLINE=93013022; PubMed=1398126;
 RA Gruber H., Goetlink S.D., Kirk D.L., Schmitt R.;
 RT "The nitrate reductase-encoding gene of Volvox carterii: map location,
 RT sequence and induction kinetics";
 RL Gene 120:75-83(1992).
 CC -1- FUNCTION: Nitrate reductase is a key enzyme involved in the first
 CC step of nitrate assimilation in plants, fungi and bacteria.
 CC -1- CATALYTIC ACTIVITY: Nitrite + NAD(+) + H(2)O = nitrate + NADH.
 CC -1- COPOLYMER: EACH SUBUNIT OF THE ENZYME CONTAINS 1 EQUIVALENT OF FAD,
 CC HEME IRON, AND POLYDENOM-PTERIN AS PROSTHETIC GROUPS. THE HEME
 CC GROUP IS CALLED CYTOCHROME B-557.
 CC -1- SUBUNIT: Homodimer (By similarity).
 CC -1- INDUCTION: By nitrate.
 CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOTERIN OXIDOREDUCTASES IN THE
 CC N-TERMINAL DOMAIN.


```

CC -1- SIMILARITY: Contains 1 cytochrome b5 heme-binding domain.
CC -1- SIMILARITY: TO FAD/NAD-BINDING CYTOCHROME REDUCTASES IN THE
CC C-TERMINAL DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X64136; CA45497.1; -.
DR PIR; JCI422; JCI422.
DR HSSP; P04166; 1BSM.
DR InterPro; IPR001199; Cyt_B5.
DR InterPro; IPR001834; Cyt_B5_reductase.
DR InterPro; IPR000572; Euk_Mb_oxred.
DR InterPro; IPR001709; FPN_Cyt_reductase.
DR InterPro; IPR005066; Mo-co_dimer.
DR InterPro; IPR001433; Oxred_FAD/NAD(P).
DR Pfam; PF00970; FAD_binding_6; 1.
DR Pfam; PF00173; heme_1; 1.
DR Pfam; PF03404; Mo-co_dimer; 1.
DR Pfam; PF00175; NAD_binding_1; 1.
DR Pfam; PF00174; oxidored_molzb; 1.
DR PRINTS; PR00406; CYTB5RDTASE.
DR PRINTS; PR00363; CYTOCHROME_B5.
DR PRINTS; PR00407; EMOPTERIN.
DR PRINTS; PR00371; FPNCR.
DR ProDom; PD000612; Cyt_B5; 1.
DR PROSITE; PS00191; CYTOCHROME_B5_1; 1.
DR PROSITE; PS00255; CYTOCHROME_B5_2; 1.
DR PROSITE; PS00559; MOLYBDOTERIN_EUK; 1.
DR Oxidoreductase; Flavoprotein; FAD; NAD; Heme; Molybdenum;
DR Nitrate assimilation.
FT METAL 139 139 MOLYBDENUM-PTERIN (POTENTIAL).
FT METAL 193 193 MOLYBDENUM-PTERIN (POTENTIAL).
FT DISULFID 376 376 INTERCHAIN (POTENTIAL).
FT METAL 532 532 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT METAL 555 555 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
SQ SEQUENCE 864 AA; 96402 MW; 499529652CDD1C7 CRC64;

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
Db 478 AAAAPAAP 486

RESULT 5
RLAL_LEIPE STANDARD; PRT; 107 AA.
AC 046313;
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S acidic ribosomal protein p1.
OS Leishmania peruviana.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5681;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOM/PE/84/LC26;
RA de los Santos M., Carrillo C., Panebra A., Montoya Y.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
CC PROTEIN SYNTHESIS.
CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
CC SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1ab-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF045249; AAC02701.1; -.
DR InterPro; IPR001813; 60S_ribosomal.
DR InterPro; IPR001859; Ribosomal_P2.
DR Pfam; PF00428; 60S_ribosomal; 1.
DR PRINTS; PR00456; RIBOSOMAL_P2.
DR Ribosomal protein.
SQ SEQUENCE 107 AA; 10896 MW; 9117DEC5E490A071 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 107;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAPA 126
Db 71 SAAAPAPA 78

RESULT 6
RLA3_SCHPO STANDARD; PRT; 110 AA.
AC P17477;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein p1-alpha 3 (A3).
GN P1A3 OR SPBC3B9.13C.
OS Schizosaccharomyces pombe (Pission Yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90220620; PubMed=2325655;
RA Beltrame M., Bianchi M.B.;
RT "A gene family for acidic ribosomal proteins in Schizosaccharomycetes
RT pombe: two essential and two nonessential genes.";
RL Mol. Cell. Biol. 10:2341-2348(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hildalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle B., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moute S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckert G., Aert R., Robben J., Gymnopoulos B.,
RA Weltyans I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gebel C., Fuchs M., Fritz C., Holzer B., Mosati D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Reger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano Z., Gloux S., Lelaire V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Kochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,

```

RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Barrell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
 CC PROTEIN SYNTHESIS.
 CC -1- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
 CC SUBUNIT.
 CC -1- MISCELLANEOUS: YEASTS CONTAIN 4 INDIVIDUAL SMALL RIBOSOMAL A
 CC PROTEINS (RPA) WHICH CAN BE CLASSIFIED INTO TWO COUPLES OF SIMILAR
 CC BUT NOT IDENTICAL SEQUENCES. EACH COUPLE IS DISTINCTLY RELATED TO
 CC ONE OF THE TWO A PROTEINS PRESENT IN MULTICELLULAR ORGANISMS.
 CC -1- MISCELLANEOUS: RPA3 AND RPA4 ARE ESSENTIAL FOR CELL SURVIVAL.
 CC WHEREAS RPA1 AND RPA2 ARE NOT.
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M31319; AAA35336.1; -;
 CC EMBL; AL022070; CA17793.1; -;
 CC PIR; C34715; R6BYP3.
 CC GeneDB_Sprobe; SPBC3B9.13c; -;
 CC InterPro; IPR001813; 609_Ribosomal.
 CC Pfam; PF00428; 609_Ribosomal; 1.
 CC KMW Ribosomal protein; Phosphorylation; Multigene family.
 CC SQ SEQUENCE 110 AA; 11171 MW; 1734AC9779F5A891 CRC64;
 CC -----
 CC Query Match 3.7%; Score 8; DB 1; Length 110;
 CC Best Local Similarity 100.0%; Pred. No. 2.3;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 119 SAAAPRA 126
 CC DB 64 SAAAPRA 71
 CC -----
 CC RESULT 7
 CC ID RL7 CHLPN STANDARD; PRT; 128 AA.
 CC AC 0925A1; 09J070;
 CC DT 30-MAY-2000 (Rel. 39, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
 CC DE 50S ribosomal protein L7/L12.
 CC GN RPL7 OR RL7 OR CPN080 OR CP0695.
 CC OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 CC OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 CC NCBI_TaxID=83558;
 CC RX MEDLINE=99206606; PubMed=10192388;
 CC RA Kaimen S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 CC Olinger U., Grimwood J., Davis R.W., Stephens R.S.;
 CC RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 CC RL Nat. Genet. 21:385-389(1999).
 CC [2]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=AR39;
 CC RX MEDLINE=20150255; PubMed=10684935;
 CC RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 CC White O., Hickey E.K., Peterson J., Uetrich T., Berry K., Bass S.,
 CC Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 CC Rahn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 CC Bisen J., Fraser C.M.;
 CC RT "Genome sequences of Chlamydia trachomatis Mohn and Chlamydia
 CC pneumoniae AK39.";

RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shiral M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWT029 from USA.";
 CC Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: SEEMS TO BE THE BINDING SITE FOR SEVERAL OF THE FACTORS
 CC INVOLVED IN PROTEIN SYNTHESIS AND APPEARS TO BE ESSENTIAL FOR
 CC ACCURATE TRANSLATION (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE001593; AAD18233.1; -;
 CC EMBL; AE002228; AAF38503.1; -;
 CC EMBL; AP002545; BAA98290.1; -;
 CC PIR; C72122; C72122.
 CC PIR; H66500; H66500.
 CC HSP; P02392; 1CTP.
 CC PHCI-2DPAGE; Q929A1; -;
 CC TIGR; CP0695; -;
 CC HAMAP; MF_00368; -; 1.
 CC InterPro; IPR000206; Ribosomal_L12.
 CC Pfam; PF00542; Ribosomal_L12; 1.
 CC ProDom; PD001326; Ribosomal_L12; 1.
 CC TIGRPAme; TIGR00855; L12; 1.
 CC KMW Ribosomal protein; Complete proteome.
 CC FT INIT MET 0
 CC BY SIMILARITY.
 CC SQ SEQUENCE 128 AA; 13461 MW; 4B2F17A85B057CC CRC64;
 CC -----
 CC Query Match 3.7%; Score 8; DB 1; Length 128;
 CC Best Local Similarity 100.0%; Pred. No. 2.6;
 CC Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 171 EAPVAAP 178
 CC DB 49 EAPVAAP 56
 CC -----
 CC RESULT 8
 CC ID CUP5 GALME STANDARD; PRT; 353 AA.
 CC AC 024958;
 CC DT 15-DEC-1998 (Rel. 37, Created)
 CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
 CC DT 15-JUL-1999 (Rel. 38, Last annotation update)
 CC DE Pupal cuticle protein PCP52 precursor (GMPCP52).
 CC GN PCP52.
 CC OS Galleria mellonella (Wax moth).
 CC OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditytaria; Pyraloidea;
 CC OC Pyralidae; Galleriinae; Galleria.
 CC NCBI_TaxID=7137;
 CC RX [1]
 CC RP SEQUENCE FROM N.A.
 CC RC TISSUE=Epidermis;
 CC RX MEDLINE=95291282; PubMed=7773255;
 CC RA Kolberg U., Obermair B., Hirsch H., Kelber G., Wolbert P.;
 CC RT "Expression cloning and characterization of a pupal cuticle protein
 CC cDNA of Galleria mellonella L.";
 CC RT Insect Biochem. Mol. Biol. 25:355-363(1995).
 CC -1- FUNCTION: COMPONENT OF THE CUTICLE OF THE PUPA OF GALLERIA
 CC MELLONELLA.

CC -1- DEVELOPMENTAL STAGE: MAXIMAL EXPRESSION DURING THE FIRST DAY AFTER
CC PUPAL ECDYSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>
CC or send an email to license@ebi.ac.uk).

DR EMBL; X77514; CAAS4650.1; -.
DR PIR; S41958; S41958.
KW Structural protein; Cuticle; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 353 PUPAL CUTICLE PROTEIN PCP52.
FT DOMAIN 235 243 POLY-ALA.
SQ SEQUENCE 353 AA; 35818 MW; C3A35B04EC049172 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 353;
Best Local Similarity 100.0%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPPA 126
Db 236 SAAAPPA 243

RESULT 9
P53_MESAU STANDARD; PRT; 396 AA.
AC Q00366; P97276;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Cellular tumor antigen p53 (Tumor suppressor p53).
GN TP53.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Syrian; TISSUE=Kidney;
RX MEDLINE=92210007; Pubmed=155573;
RA Legros Y., McIntyre P., Soussi T.;
RL "The cDNA cloning and immunological characterization of hamster p53."
RN Gene 112:247-250(1992).
RP SEQUENCE FROM N.A.
RA Hou E.W., Wiseman R.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Acts as a tumor suppressor in many tumor types; induces
CC growth arrest or apoptosis depending on the physiological
CC circumstances and cell type. Involved in cell cycle regulation as
CC a trans-activator that acts to negatively regulate cell division
CC by controlling a set of genes required for this process. One of
CC the activated genes is an inhibitor of cyclin-dependent kinases.
CC Apoptosis induction seems to be mediated either by stimulation of
CC BAX and BAK antigen expression, or by repression of Bcl-2
CC expression.
CC -1- SUBUNIT: Binds DNA as a homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Nucleus.
CC -1- DISSEMIN: p53 is found in increased amounts in a wide variety
CC of transformed cells. p53 is frequently mutated or inactivated
CC in many types of cancer.
CC -1- SIMILARITY: BELONGS TO THE P53 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announce/>
CC or send an email to license@ebi.ac.uk).

DR EMBL; M75144; AAA37085.1; -.
DR EMBL; U07182; AAB41344.1; -.
DR PIR; JH0633; JH0633.
DR HSSP; P04637; TRUP.
DR InterPro; IPR02117; P53.
DR Pfam; PF00870; P53.1.
DR PRINTS; PR00386; P53SUPPRESSR.
DR ProDom; PD002681; P53; 1.
DR PROSITE; PS00348; P53; 1.
KW Anti-oncogene; DNA-binding; Transcription regulation; Activator;
KW Nuclear protein; Phosphorylation; Apoptosis.
FT DOMAIN 1 45 TRANSCRIPTION ACTIVATION (ACIDIC).
FT DNA_BIND 105 295 BY SIMILARITY.
FT DOMAIN 328 359 OLIGOMERIZATION.
FT DOMAIN 371 390 BASIC (REPRESSION OF DNA-BINDING).
FT DOMAIN 314 326 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 15 15 PHOSPHORYLATION (BY PRPK) (BY
FT MOD_RES 395 395 PHOSPHORYLATION (BY SIMILARITY).
FT CONFLICT 188 188 G -> S (IN REF. 2).
SQ SEQUENCE 396 AA; 43631 MW; 906EF02568099BE3 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 396;
Best Local Similarity 100.0%; Pred. No. 6.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
Db 66 AAAAPAP 73

RESULT 10
ATP2_CHLRE STANDARD; PRT; 574 AA.
ID ATP2_CHLRE
AC P384B2;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE ATP synthase beta chain, mitochondrial precursor (EC 3.6.3.14).
GN ATP2.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9235336; Pubmed=1386535;
RA Franzen L.-G., Falk G.;
RL "Nucleotide sequence of cDNA clones encoding the beta subunit of
RL mitochondrial ATP synthase from the green alga Chlamydomonas
RL reinhardtii: the precursor protein encoded by the cDNA contains both
RL an N-terminal presequence and a C-terminal extension.";
RN Plant Mol. Biol. 19:771-780(1992).
RP STRUCTURE BY NMR OF 1-26.
RX MEDLINE=9632653; Pubmed=8706917;
RA Lancelotti J.-M., Gane P., Bouchayer E., Bailly I., Arlaud G.J.,
RA Jacquot J.-P.;
RL "NMR structures of a mitochondrial transmembrane peptide from the green
RL alga Chlamydomonas reinhardtii.";
RL PDBS Lett. 391:203-208(1996).
CC -1- FUNCTION: PRODUCES ATP FROM ADP IN THE PRESENCE OF A PROTON
CC GRADIENT ACROSS THE MEMBRANE. THE BETA CHAIN IS THE CATALYTIC
CC SUBUNIT.
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate +
CC H(+) (out).
CC -1- SUBUNIT: F-TYPE APPARUS HAVE 2 COMPONENTS. CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), Epsilon(1). CF(0)

```

CC      HAS THREE MAIN SUBUNITS: A, B AND C.
CC      -1- SUBCELLULAR LOCATION: Mitochondrial.
CC      -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; X61624; CAA43808.1; -.
CC      PIR; S23530; S23530.
CC      HSSP; P00829; 1BMF.
CC      InterPro; IPR003593; AAA_ATPase.
CC      InterPro; IPR000793; ATPase_a/bc.
CC      InterPro; IPR000194; ATPase_a/bc.
CC      InterPro; IPR004100; ATPase_a/bn.
CC      InterPro; IPR005722; F1_ATPase_beta.
CC      Pfam; PF00006; ATP-synt_ab_C; 1.
CC      Pfam; PF00306; ATP-synt_ab_C; 1.
CC      Pfam; PF02874; ATP-synt_ab_N; 1.
CC      SMART; SM00382; AAA; 1.
CC      TIGRfam; TIGR01039; atpD; 1.
CC      PROSITE; PS00152; ATPASE ALPHA BETA; 1.
CC      ATP synthase; CF(1); Hydrogen ion transport;
CC      Hydrolyase; ATP-binding; Mitochondrion; Transil.
CC      TRANSIT 1 26 MITOCHONDRION (POTENTIAL).
CC      CHAIN 27 574 ATP SYNTHASE BETA CHAIN.
CC      NP BIND 183 190 ATP (BY SIMILARITY).
CC      SEQUENCE 574 AA; 61821 MW; 2283C6C6D18FBCFE CRC64;

Query Match      3.7%; Score 8; DB 1; Length 574;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      130 AAEPAANA 137
DB      26 AAEPAANA 33

RESULT 11
MSPA TREMA STANDARD; PRT; 575 AA.
AC 092413;
ID MSPA TREMA STANDARD; PRT; 575 AA.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Major outer membrane protein mspa precursor (Major sheath protein).
GN MSPA.
OS Treponema maltophilum.
OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=51150;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 20-34.
RC STRAIN=ATCC 51939 / BR;
RX MEDLINE=99121045; PubMed=9922270;
RA Heuner K., Choi B.-K., Schade R., Moter A., Otto A., Goebel U.B.;
RT Cloning and characterization of a gene (mspa) encoding the major
RT sheath protein of Treponema maltophilum ATCC 51939 (T)";
RT J. Bacteriol. 181:1025-1029 (1999).
RN [2]
RN SUBCELLULAR LOCATION.
RX MEDLINE=21213882; PubMed=1131333;
RA Heuner K., Meltzer U., Choi B.-K., Goebel U.B.;
RT "Outer sheath associated proteins of the oral spirochete Treponema
RT maltophilum.";
RL FEMS Microbiol. Lett. 197:187-193 (2001).
CC -1- FUNCTION: Major component of the outer membrane sheath.
CC -1- SUBCELLULAR LOCATION: Outer membrane-associated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; Y17800; CAA76862.1; -.
CC      PIR; S17800; CAA76862.1; -.
CC      HSSP; P00047; STROIDPINGER.
CC      InterPro; IPR001628; znf_C4steroid.
CC      Pfam; PF00104; hormone_rec; 1.
CC      Pfam; PF00105; zf-C4; 1.
CC      PRINTS; PR00398; STRDHOMONER.
CC      PRINTS; PR00047; STROIDPINGER.
CC      ProDom; PD000035; znf_C4steroid; 1.
CC      SMART; SM00430; HOL1; 1.
CC      SMART; SM00399; ZNF C4; 1.
CC      PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.

Query Match      3.7%; Score 8; DB 1; Length 575;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      155 AAEPAEA 162
DB      19 AAEPAEA 26

RESULT 12
ECR HELVI STANDARD; PRT; 576 AA.
AC 018473;
ID ECR HELVI STANDARD; PRT; 576 AA.
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ecdysone receptor (Ecdy/steroid receptor) (20-hydroxy-ecdysone
DE receptor) (20R receptor) (HVECR).
GN ECR OR NR1H1.
OS Heliothis virescens (Noctuid moth) (Owl moth).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Lepidoptera; Glossata; Ditryaia; Noctuoidea;
OC Noctuidae; Heliothinae; Heliothis.
OX NCBI_TaxID=7102;
RN [1]
RN SEQUENCE FROM N.A.
RN STRAIN=BR;
RX MEDLINE=99457718; PubMed=10528411;
RA Martinez A., Scanlon D., Gross B., Petras S.C., Ball S.R.,
RA Greenland A.V., Windas J., Pong O., Broad P., Jepson I.;
RT "Transcriptional activation of the cloned Heliothis virescens
RT (Lepidoptera) ecdysone receptor (HVECR) by muristeroneA.";
RT Insect Biochem. Mol. Biol. 29:915-930 (1999).
CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE
CC ELEMENTS (ECRES) (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR1
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC      EMBL; Y09009; CAA70212.1; -.
CC      HSSP; P20393; 1A6Y.
CC      InterPro; IPR000536; Hormone_rec_1lg.
CC      InterPro; IPR001723; Strdhm receptor.
CC      InterPro; IPR001628; znf_C4steroid.
CC      Pfam; PF00104; hormone_rec; 1.
CC      Pfam; PF00105; zf-C4; 1.
CC      PRINTS; PR00398; STRDHOMONER.
CC      PRINTS; PR00047; STROIDPINGER.
CC      ProDom; PD000035; znf_C4steroid; 1.
CC      SMART; SM00430; HOL1; 1.
CC      SMART; SM00399; ZNF C4; 1.
CC      PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.

```

KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
 KM Zinc-finger.
 FT DOMAIN 1 162 MODULATING (POTENTIAL).
 FT DNA BIND 163 228 NUCLEAR RECEPTOR-TYPE.
 FT ZN_FING 163 228 C4-TYPE.
 FT ZN_FING 199 223 C4-TYPE.
 FT DOMAIN 326 545 HORMONE-BINDING (POTENTIAL).
 SQ SEQUENCE 576 AA; 64638 MW; D3EE787BF263A8 CRC64;
 Query March 3.7%; Score 8; DB 1; Length 576;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 140 TATPVAE 147
 DB 552 TATPVAE 559
 RESULT 13
 ID KELC DROME STANDARD; PRT; 1477 AA.
 AC 004653; 09VIA2;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Ring canal kelch protein (Contains: Kelch short protein).
 GN KEL OR CG210.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pelecyota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo.
 RX MEDLINE=93201592; PubMed=8453663;
 RA Xue F., Cooley L.;
 RT "Kelch encodes a component of intercellular bridges in Drosophila egg
 chambers."
 RL Cell 72:681-693(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Anagnostis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballow R.M., Baer A., Baxendale J., Bayraktarglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brinkstein P., Brothier P.,
 RA Butts K.C., Buser D.A., Butler H., Cadieu E., Canter A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablo S., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertile S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodde A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson S.D., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefel F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson W., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Sytkae R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [3]
 RP CHARACTERIZATION.
 RC TISSUE=Embryo;
 RX MEDLINE=97236487; PubMed=9118811;
 RA Robinson D.N., Cooley L.;
 RT "Examination of the function of two kelch proteins generated by stop
 codon suppression."
 RL Development 124:1405-1417(1997).
 CC -1- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF
 CYTOPLASM BETWEEN CELLS. MAY BE INVOLVED IN THE REGULATION OF
 CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
 CC BINDS ACTIN.
 CC -1- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING
 CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK
 EPITHELIA.
 CC -1- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE
 TESTIS, OVARECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND
 IMAGINAL DISKS. KELCH ORP1 IS THE PREDOMINANT PROTEIN AND IS ALSO
 EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF
 FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.
 CC -1- DEVELOPMENTAL STAGE: LARVAE; PUPAE AND ADULTS.
 CC -1- SIMILARITY: Contains 6 BTB/POZ domain.
 CC -1- CAUTION: REF.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 gene model prediction.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1eb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL; L08483; AAA53471.1; -
 DR EMBL; L08483; AAA53472.2; -
 DR EMBL; AE003657; AAF53651.1; ALT_SEQ.
 DR HSSP; Q05516; 1C63.
 DR FLYBase; FBgn0001301; Kel.
 DR GO; GO:0007292; P:ogenesis; IMP.
 DR GO; GO:0007301; P:ring canal formation; IDA.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR006511; Kelch.
 DR InterPro; IPR006512; Kelch_rep.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF01344; Kelch; 6.
 DR PRINTS; PR00501; KELCHREPEAT.
 DR SMART; SM00225; BTB; 1.
 DR SMART; SM00612; Kelch; 6.
 DR PROSITE; PS50057; BTB; 1.
 KW Cytokeleton; Actin-binding; Selenum; Selenocysteine; Kelch repeat;
 KM Repeat.
 FT CHAIN 1 1477 KELCH PROTEIN.
 FT CHAIN 1 689 KELCH SHORT PROTEIN.
 FT REPEAT 157 223 BTB.
 FT REPEAT 404 449 KELCH 1.
 FT REPEAT 450 496 KELCH 2.
 FT REPEAT 498 543 KELCH 3.
 FT REPEAT 545 592 KELCH 4.
 FT REPEAT 594 639 KELCH 5.
 FT REPEAT 641 687 KELCH 6.
 FT DOMAIN 18 28 ASN-RICH.
 FT DOMAIN 29 87 GLN-RICH.

```

FT DOMAIN 29 36 POLY-GLN.
FT DOMAIN 78 83 POLY-GLN.
FT SE CYS 690 690 PROBABLE.
FT CONFLICT 493 493 V -> A (IN REF. 1).
FT CONFLICT 596 596 A -> R (IN REF. 1).
FT CONFLICT 824 824 P -> L (IN REF. 1).
FT CONFLICT 858 858 G -> D (IN REF. 1).
FT CONFLICT 1083 1083 A -> R (IN REF. 1).
FT CONFLICT 1086 1086 A -> G (IN REF. 1).
SQ SEQUENCE 1477 AA; 160086 MW; 4851BEA9D9DBA7 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 1477;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 120 AAAAPAP 127
Db 740 AAAAPAP 747

RESULT 14
NTC2 MOUSE STANDARD; PRT; 2470 AA.
ID NTC2_MOUSE Q06008; Q060941;
AC Q35516; Q06008; Q060941;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Neogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
DE B)
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Thymus;
RA Hamada Y., Higuchi M., Tsujimoto Y.;
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene."
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 316-1518 FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
RA MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.;
RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
RT wide variety of tissues."
RL Exp. Cell Res. 204:364-372 (1993).
RN [3]
RP SEQUENCE OF 1765-2153 FROM N.A.
RC MEDLINE=97075110; PubMed=8917536;
RA Milner L.A., Bigas A., Kopan R., Brashem-Stein C., Bernstein I.D.,
RA Martin D.I.;
RT "Inhibition of granulocytic differentiation by mNotch1."
RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019 (1996).
RN [4]
RP FUNCTION.
RA MEDLINE=99396706; PubMed=10393120;
RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
RA Tsujimoto Y.;
RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
RT embryonic lethality."
RL Development 126:3415-3424 (1999).
RN [5]
RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
RA MEDLINE=95333893; PubMed=7609614;
RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
RT "Differential expression of Notch1 and Notch2 in developing and adult
RT mouse brain."
RL Brain Res. Mol. Brain Res. 29:263-272 (1995).
RN [6]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.

```

```

RX MEDLINE=21523956; PubMed=11518718;
RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
RT "Murine notch homologs (N1-4) undergo presenilin-dependent
RT proteolysis."
RL J. Biol. Chem. 276:40268-40273 (2001).
RN [7]
RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
RX MEDLINE=21374376; PubMed=11459941;
RA Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
RT "Conservation of the biochemical mechanisms of signal transduction
RT among mammalian Notch family members."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031 (2001).
CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
CC Jagged1, Jagged2 and Deltal to regulate cell-fate determination.
CC Upon ligand activation through the released notch intracellular
CC domain (NICD) it forms a transcriptional activator complex with
CC RBP-J kappa and activates genes of the enhancer of split locus.
CC Affects the implementation of differentiation, proliferation and
CC apoptotic programs (By similarity). May play an essential role in
CC postimplantation development, probably in some aspect of cell
CC specification and/or differentiation.
CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TW) and a N-
CC terminal fragment N(EC) which are probably linked by disulfide
CC bonds.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
CC proteolytical processing NICD is translocated to the nucleus.
CC -1- ALTERNATIVE PRODUCTS:
CC Name=1;
CC Name=2;
CC Name=3;
CC Name=4;
CC Name=5;
CC Name=6;
CC Name=7;
CC Name=8;
CC Name=9;
CC Name=10;
CC Name=11;
CC Name=12;
CC Name=13;
CC Name=14;
CC Name=15;
CC Name=16;
CC Name=17;
CC Name=18;
CC Name=19;
CC Name=20;
CC Name=21;
CC Name=22;
CC Name=23;
CC Name=24;
CC Name=25;
CC Name=26;
CC Name=27;
CC Name=28;
CC Name=29;
CC Name=30;
CC Name=31;
CC Name=32;
CC Name=33;
CC Name=34;
CC Name=35;
CC Name=36;
CC Name=37;
CC Name=38;
CC Name=39;
CC Name=40;
CC Name=41;
CC Name=42;
CC Name=43;
CC Name=44;
CC Name=45;
CC Name=46;
CC Name=47;
CC Name=48;
CC Name=49;
CC Name=50;
CC Name=51;
CC Name=52;
CC Name=53;
CC Name=54;
CC Name=55;
CC Name=56;
CC Name=57;
CC Name=58;
CC Name=59;
CC Name=60;
CC Name=61;
CC Name=62;
CC Name=63;
CC Name=64;
CC Name=65;
CC Name=66;
CC Name=67;
CC Name=68;
CC Name=69;
CC Name=70;
CC Name=71;
CC Name=72;
CC Name=73;
CC Name=74;
CC Name=75;
CC Name=76;
CC Name=77;
CC Name=78;
CC Name=79;
CC Name=80;
CC Name=81;
CC Name=82;
CC Name=83;
CC Name=84;
CC Name=85;
CC Name=86;
CC Name=87;
CC Name=88;
CC Name=89;
CC Name=90;
CC Name=91;
CC Name=92;
CC Name=93;
CC Name=94;
CC Name=95;
CC Name=96;
CC Name=97;
CC Name=98;
CC Name=99;
CC Name=100;
CC Name=101;
CC Name=102;
CC Name=103;
CC Name=104;
CC Name=105;
CC Name=106;
CC Name=107;
CC Name=108;
CC Name=109;
CC Name=110;
CC Name=111;
CC Name=112;
CC Name=113;
CC Name=114;
CC Name=115;
CC Name=116;
CC Name=117;
CC Name=118;
CC Name=119;
CC Name=120;
CC Name=121;
CC Name=122;
CC Name=123;
CC Name=124;
CC Name=125;
CC Name=126;
CC Name=127;
CC Name=128;
CC Name=129;
CC Name=130;
CC Name=131;
CC Name=132;
CC Name=133;
CC Name=134;
CC Name=135;
CC Name=136;
CC Name=137;
CC Name=138;
CC Name=139;
CC Name=140;
CC Name=141;
CC Name=142;
CC Name=143;
CC Name=144;
CC Name=145;
CC Name=146;
CC Name=147;
CC Name=148;
CC Name=149;
CC Name=150;
CC Name=151;
CC Name=152;
CC Name=153;
CC Name=154;
CC Name=155;
CC Name=156;
CC Name=157;
CC Name=158;
CC Name=159;
CC Name=160;
CC Name=161;
CC Name=162;
CC Name=163;
CC Name=164;
CC Name=165;
CC Name=166;
CC Name=167;
CC Name=168;
CC Name=169;
CC Name=170;
CC Name=171;
CC Name=172;
CC Name=173;
CC Name=174;
CC Name=175;
CC Name=176;
CC Name=177;
CC Name=178;
CC Name=179;
CC Name=180;
CC Name=181;
CC Name=182;
CC Name=183;
CC Name=184;
CC Name=185;
CC Name=186;
CC Name=187;
CC Name=188;
CC Name=189;
CC Name=190;
CC Name=191;
CC Name=192;
CC Name=193;
CC Name=194;
CC Name=195;
CC Name=196;
CC Name=197;
CC Name=198;
CC Name=199;
CC Name=200;
CC Name=201;
CC Name=202;
CC Name=203;
CC Name=204;
CC Name=205;
CC Name=206;
CC Name=207;
CC Name=208;
CC Name=209;
CC Name=210;
CC Name=211;
CC Name=212;
CC Name=213;
CC Name=214;
CC Name=215;
CC Name=216;
CC Name=217;
CC Name=218;
CC Name=219;
CC Name=220;
CC Name=221;
CC Name=222;
CC Name=223;
CC Name=224;
CC Name=225;
CC Name=226;
CC Name=227;
CC Name=228;
CC Name=229;
CC Name=230;
CC Name=231;
CC Name=232;
CC Name=233;
CC Name=234;
CC Name=235;
CC Name=236;
CC Name=237;
CC Name=238;
CC Name=239;
CC Name=240;
CC Name=241;
CC Name=242;
CC Name=243;
CC Name=244;
CC Name=245;
CC Name=246;
CC Name=247;
CC Name=248;
CC Name=249;
CC Name=250;
CC Name=251;
CC Name=252;
CC Name=253;
CC Name=254;
CC Name=255;
CC Name=256;
CC Name=257;
CC Name=258;
CC Name=259;
CC Name=260;
CC Name=261;
CC Name=262;
CC Name=263;
CC Name=264;
CC Name=265;
CC Name=266;
CC Name=267;
CC Name=268;
CC Name=269;
CC Name=270;
CC Name=271;
CC Name=272;
CC Name=273;
CC Name=274;
CC Name=275;
CC Name=276;
CC Name=277;
CC Name=278;
CC Name=279;
CC Name=280;
CC Name=281;
CC Name=282;
CC Name=283;
CC Name=284;
CC Name=285;
CC Name=286;
CC Name=287;
CC Name=288;
CC Name=289;
CC Name=290;
CC Name=291;
CC Name=292;
CC Name=293;
CC Name=294;
CC Name=295;
CC Name=296;
CC Name=297;
CC Name=298;
CC Name=299;
CC Name=300;
CC Name=301;
CC Name=302;
CC Name=303;
CC Name=304;
CC Name=305;
CC Name=306;
CC Name=307;
CC Name=308;
CC Name=309;
CC Name=310;
CC Name=311;
CC Name=312;
CC Name=313;
CC Name=314;
CC Name=315;
CC Name=316;
CC Name=317;
CC Name=318;
CC Name=319;
CC Name=320;
CC Name=321;
CC Name=322;
CC Name=323;
CC Name=324;
CC Name=325;
CC Name=326;
CC Name=327;
CC Name=328;
CC Name=329;
CC Name=330;
CC Name=331;
CC Name=332;
CC Name=333;
CC Name=334;
CC Name=335;
CC Name=336;
CC Name=337;
CC Name=338;
CC Name=339;
CC Name=340;
CC Name=341;
CC Name=342;
CC Name=343;
CC Name=344;
CC Name=345;
CC Name=346;
CC Name=347;
CC Name=348;
CC Name=349;
CC Name=350;
CC Name=351;
CC Name=352;
CC Name=353;
CC Name=354;
CC Name=355;
CC Name=356;
CC Name=357;
CC Name=358;
CC Name=359;
CC Name=360;
CC Name=361;
CC Name=362;
CC Name=363;
CC Name=364;
CC Name=365;
CC Name=366;
CC Name=367;
CC Name=368;
CC Name=369;
CC Name=370;
CC Name=371;
CC Name=372;
CC Name=373;
CC Name=374;
CC Name=375;
CC Name=376;
CC Name=377;
CC Name=378;
CC Name=379;
CC Name=380;
CC Name=381;
CC Name=382;
CC Name=383;
CC Name=384;
CC Name=385;
CC Name=386;
CC Name=387;
CC Name=388;
CC Name=389;
CC Name=390;
CC Name=391;
CC Name=392;
CC Name=393;
CC Name=394;
CC Name=395;
CC Name=396;
CC Name=397;
CC Name=398;
CC Name=399;
CC Name=400;
CC Name=401;
CC Name=402;
CC Name=403;
CC Name=404;
CC Name=405;
CC Name=406;
CC Name=407;
CC Name=408;
CC Name=409;
CC Name=410;
CC Name=411;
CC Name=412;
CC Name=413;
CC Name=414;
CC Name=415;
CC Name=416;
CC Name=417;
CC Name=418;
CC Name=419;
CC Name=420;
CC Name=421;
CC Name=422;
CC Name=423;
CC Name=424;
CC Name=425;
CC Name=426;
CC Name=427;
CC Name=428;
CC Name=429;
CC Name=430;
CC Name=431;
CC Name=432;
CC Name=433;
CC Name=434;
CC Name=435;
CC Name=436;
CC Name=437;
CC Name=438;
CC Name=439;
CC Name=440;
CC Name=441;
CC Name=442;
CC Name=443;
CC Name=444;
CC Name=445;
CC Name=446;
CC Name=447;
CC Name=448;
CC Name=449;
CC Name=450;
CC Name=451;
CC Name=452;
CC Name=453;
CC Name=454;
CC Name=455;
CC Name=456;
CC Name=457;
CC Name=458;
CC Name=459;
CC Name=460;
CC Name=461;
CC Name=462;
CC Name=463;
CC Name=464;
CC Name=465;
CC Name=466;
CC Name=467;
CC Name=468;
CC Name=469;
CC Name=470;
CC Name=471;
CC Name=472;
CC Name=473;
CC Name=474;
CC Name=475;
CC Name=476;
CC Name=477;
CC Name=478;
CC Name=479;
CC Name=480;
CC Name=481;
CC Name=482;
CC Name=483;
CC Name=484;
CC Name=485;
CC Name=486;
CC Name=487;
CC Name=488;
CC Name=489;
CC Name=490;
CC Name=491;
CC Name=492;
CC Name=493;
CC Name=494;
CC Name=495;
CC Name=496;
CC Name=497;
CC Name=498;
CC Name=499;
CC Name=500;
CC Name=501;
CC Name=502;
CC Name=503;
CC Name=504;
CC Name=505;
CC Name=506;
CC Name=507;
CC Name=508;
CC Name=509;
CC Name=510;
CC Name=511;
CC Name=512;
CC Name=513;
CC Name=514;
CC Name=515;
CC Name=516;
CC Name=517;
CC Name=518;
CC Name=519;
CC Name=520;
CC Name=521;
CC Name=522;
CC Name=523;
CC Name=524;
CC Name=525;
CC Name=526;
CC Name=527;
CC Name=528;
CC Name=529;
CC Name=530;
CC Name=531;
CC Name=532;
CC Name=533;
CC Name=534;
CC Name=535;
CC Name=536;
CC Name=537;
CC Name=538;
CC Name=539;
CC Name=540;
CC Name=541;
CC Name=542;
CC Name=543;
CC Name=544;
CC Name=545;
CC Name=546;
CC Name=547;
CC Name=548;
CC Name=549;
CC Name=550;
CC Name=551;
CC Name=552;
CC Name=553;
CC Name=554;
CC Name=555;
CC Name=556;
CC Name=557;
CC Name=558;
CC Name=559;
CC Name=560;
CC Name=561;
CC Name=562;
CC Name=563;
CC Name=564;
CC Name=565;
CC Name=566;
CC Name=567;
CC Name=568;
CC Name=569;
CC Name=570;
CC Name=571;
CC Name=572;
CC Name=573;
CC Name=574;
CC Name=575;
CC Name=576;
CC Name=577;
CC Name=578;
CC Name=579;
CC Name=580;
CC Name=581;
CC Name=582;
CC Name=583;
CC Name=584;
CC Name=585;
CC Name=586;
CC Name=587;
CC Name=588;
CC Name=589;
CC Name=590;
CC Name=591;
CC Name=592;
CC Name=593;
CC Name=594;
CC Name=595;
CC Name=596;
CC Name=597;
CC Name=598;
CC Name=599;
CC Name=600;
CC Name=601;
CC Name=602;
CC Name=603;
CC Name=604;
CC Name=605;
CC Name=606;
CC Name=607;
CC Name=608;
CC Name=609;
CC Name=610;
CC Name=611;
CC Name=612;
CC Name=613;
CC Name=614;
CC Name=615;
CC Name=616;
CC Name=617;
CC Name=618;
CC Name=619;
CC Name=620;
CC Name=621;
CC Name=622;
CC Name=623;
CC Name=624;
CC Name=625;
CC Name=626;
CC Name=627;
CC Name=628;
CC Name=629;
CC Name=630;
CC Name=631;
CC Name=632;
CC Name=633;
CC Name=634;
CC Name=635;
CC Name=636;
CC Name=637;
CC Name=638;
CC Name=639;
CC Name=640;
CC Name=641;
CC Name=642;
CC Name=643;
CC Name=644;
CC Name=645;
CC Name=646;
CC Name=647;
CC Name=648;
CC Name=649;
CC Name=650;
CC Name=651;
CC Name=652;
CC Name=653;
CC Name=654;
CC Name=655;
CC Name=656;
CC Name=657;
CC Name=658;
CC Name=659;
CC Name=660;
CC Name=661;
CC Name=662;
CC Name=663;
CC Name=664;
CC Name=665;
CC Name=666;
CC Name=667;
CC Name=668;
CC Name=669;
CC Name=670;
CC Name=671;
CC Name=672;
CC Name=673;
CC Name=674;
CC Name=675;
CC Name=676;
CC Name=677;
CC Name=678;
CC Name=679;
CC Name=680;
CC Name=681;
CC Name=682;
CC Name=683;
CC Name=684;
CC Name=685;
CC Name=686;
CC Name=687;
CC Name=688;
CC Name=689;
CC Name=690;
CC Name=691;
CC Name=692;
CC Name=693;
CC Name=694;
CC Name=695;
CC Name=696;
CC Name=697;
CC Name=698;
CC Name=699;
CC Name=700;
CC Name=701;
CC Name=702;
CC Name=703;
CC Name=704;
CC Name=705;
CC Name=706;
CC Name=707;
CC Name=708;
CC Name=709;
CC Name=710;
CC Name=711;
CC Name=712;
CC Name=713;
CC Name=714;
CC Name=715;
CC Name=716;
CC Name=717;
CC Name=718;
CC Name=719;
CC Name=720;
CC Name=721;
CC Name=722;
CC Name=723;
CC Name=724;
CC Name=725;
CC Name=726;
CC Name=727;
CC Name=728;
CC Name=729;
CC Name=730;
CC Name=731;
CC Name=732;
CC Name=733;
CC Name=734;
CC Name=735;
CC Name=736;
CC Name=737;
CC Name=738;
CC Name=739;
CC Name=740;
CC Name=741;
CC Name=742;
CC Name=743;
CC Name=744;
CC Name=745;
CC Name=746;
CC Name=747;
CC Name=748;
CC Name=749;
CC Name=750;
CC Name=751;
CC Name=752;
CC Name=753;
CC Name=754;
CC Name=755;
CC Name=756;
CC Name=757;
CC Name=758;
CC Name=759;
CC Name=760;
CC Name=761;
CC Name=762;
CC Name=763;
CC Name=764;
CC Name=765;
CC Name=766;
CC Name=767;
CC Name=768;
CC Name=769;
CC Name=770;
CC Name=771;
CC Name=772;
CC Name=773;
CC Name=774;
CC Name=775;
CC Name=776;
CC Name=777;
CC Name=778;
CC Name=779;
CC Name=780;
CC Name=781;
CC Name=782;
CC Name=783;
CC Name=784;
CC Name=785;
CC Name=786;
CC Name=787;
CC Name=788;
CC Name=789;
CC Name=790;
CC Name=791;
CC Name=792;
CC Name=793;
CC Name=794;
CC Name=795;
CC Name=796;
CC Name=797;
CC Name=798;
CC Name=799;
CC Name=800;
CC Name=801;
CC Name=802;
CC Name=803;
CC Name=804;
CC Name=805;
CC Name=806;
CC Name=807;
CC Name=808;
CC Name=809;
CC Name=810;
CC Name=811;
CC Name=812;
CC Name=813;
CC Name=814;
CC Name=815;
CC Name=816;
CC Name=817;
CC Name=818;
CC Name=819;
CC Name=820;
CC Name=821;
CC Name=822;
CC Name=823;
CC Name=824;
CC Name=825;
CC Name=826;
CC Name=827;
CC Name=828;
CC Name=829;
CC Name=830;
CC Name=831;
CC Name=832;
CC Name=833;
CC Name=834;
CC Name=835;
CC Name=836;
CC Name=837;
CC Name=838;
CC Name=839;
CC Name=840;
CC Name=841;
CC Name=842;
CC Name=843;
CC Name=844;
CC Name=845;
CC Name=846;
CC Name=847;
CC Name=848;
CC Name=849;
CC Name=850;
CC Name=851;
CC Name=852;
CC Name=853;
CC Name=854;
CC Name=855;
CC Name=856;
CC Name=857;
CC Name=858;
CC Name=859;
CC Name=860;
CC Name=861;
CC Name=862;
CC Name=863;
CC Name=864;
CC Name=865;
CC Name=866;
CC Name=867;
CC Name=868;
CC Name=869;
CC Name=870;
CC Name=871;
CC Name=872;
CC Name=873;
CC Name=874;
CC Name=875;
CC Name=876;
CC Name=877;
CC Name=878;
CC Name=879;
CC Name=880;
CC Name=881;
CC Name=882;
CC Name=883;
CC Name=884;
CC Name=885;
CC Name=886;
CC Name=887;
CC Name=888;
CC Name=889;
CC Name=890;
CC Name=891;
CC Name=892;
CC Name=893;
CC Name=894;
CC Name=895;
CC Name=896;
CC Name=897;
CC Name=898;
CC Name=899;
CC Name=900;
CC Name=901;
CC Name=902;
CC Name=903;
CC Name=904;
CC Name=905;
CC Name=906;
CC Name=907;
CC Name=908;
CC Name=909;
CC Name=910;
CC Name=911;
CC Name=912;
CC Name=913;
CC Name=914;
CC Name=915;
CC Name=916;
CC Name=917;
CC Name=918;
CC Name=919;
CC Name=920;
CC Name=921;
CC Name=922;
CC Name=923;
CC Name=924;
CC Name=925;
CC Name=926;
CC Name=927;
CC Name=928;
CC Name=929;
CC Name=930;
CC Name=931;
CC Name=932;
CC Name=933;
CC Name=934;
CC Name=935;
CC Name=936;
CC Name=937;
CC Name=938;
CC Name=939;
CC Name=940;
CC Name=941;
CC Name=942;
CC Name=943;
CC Name=944;
CC Name=945;
CC Name=946;
CC Name=947;
CC Name=948;
CC Name=949;
CC Name=950;
CC Name=951;
CC Name=952;
CC Name=953;
CC Name=954;
CC Name=955;
CC Name=956;
CC Name=957;
CC Name=958;
CC Name=959;
CC Name=960;
CC Name=961;
CC Name=962;
CC Name=963;
CC Name=964;
CC Name=965;
CC Name=966;
CC Name=967;
CC Name=968;
CC Name=969;
CC Name=970;
CC Name=971;
CC Name=972;
CC Name=973;
CC Name=974;
CC Name=975;
CC Name=976;
CC Name=977;
CC Name=978;
CC Name=979;
CC Name=980;
CC Name=981;
CC Name=982;
CC Name=983;
CC Name=984;
CC Name=985;
CC Name=986;
CC Name=987;
CC Name=988;
CC Name=989;
CC Name=990;
CC Name=991;
CC Name=992;
CC Name=993;
CC Name=994;
CC Name=995;
CC Name=996;
CC Name=997;
CC Name=998;
CC Name=999;
CC Name=1000;
CC Name=1001;
CC Name=1002;
CC Name=1003;
CC Name=1004;
CC Name=1005;
CC Name=1006;
CC Name=1007;
CC Name=1008;
CC Name=1009;
CC Name=1010;
CC Name=1011;
CC Name=1012;
CC Name=1013;
CC Name=1014;
CC Name=1015;
CC Name=1016;
CC Name=1017;
CC Name=1018;
CC Name=1019;
CC Name=1020;
CC Name=1021;
CC Name=1022;
CC Name=1023;
CC Name=1024;
CC Name=1025;
CC Name=1026;
CC Name=1027;
CC Name=1028;
CC Name=1029;
CC Name=1030;
CC Name=1031;
CC Name=1032;
CC Name=1033;
CC Name=1034;
CC Name=1035;
CC Name=1036;
CC Name=1037;
CC Name=1038;
CC Name=1039;
CC Name=1040;
CC Name=1041;
CC Name=1042;
CC Name=1043;
CC Name=1044;
CC Name=1045;
CC Name=1046;
CC Name=1047;
CC Name=1048;
CC Name=1049;
CC Name=1050;
CC Name=1051;
CC Name=1052;
CC Name=1053;
CC Name=1054;
CC Name=1055;
CC Name=1056;
CC Name=1057;
CC Name=1058;
CC Name=1059;
CC Name=1060;
CC Name=1061;
CC Name=1062;
CC Name=1063;
CC Name=1064;
CC Name=1065;
CC Name=1066;
CC Name=1067;
CC Name=1068;
CC Name=1069;
CC Name=1070;
CC Name=1071;
CC Name=1072;
CC Name=1073;
CC Name=1074;
CC Name=1075;
CC Name=1076;
CC Name=1077;
CC Name=1078;
CC Name=1079;
CC Name=1080;
CC Name=1081;
CC Name=1082;
CC Name=1083;
CC Name=1084;
CC Name=1085;
CC Name=1086;
CC Name=1087;
CC Name=1088;
CC Name=1089;
CC Name=1090;
CC Name=1091;
CC Name=1092;
CC Name=1093;
CC Name=1094;
CC Name=1095;
CC Name=1096;
CC Name=1097;
CC Name=1098;
CC Name=1099;
CC Name=1100;
CC Name=1101;
CC Name=1102;
CC Name=1103;
CC Name=1104;
CC Name=1105;
CC Name=1106;
CC Name=1107;
CC Name=1108;
CC Name=1109;
CC Name=1110;
CC Name=1111;
CC Name=1112;
CC Name=1113;
CC Name=1114;
CC Name=1115;
CC Name=1116;
CC Name=1117;
CC Name=1118;
CC Name=1119;
CC Name=1120;
CC Name=1121;
CC Name=1122;
CC Name=1123;
CC Name=1124;
CC Name=1125;
CC Name=1126;
CC Name=1127;
CC Name=1128;
CC Name=1129;
CC Name=1130;
CC Name=1131;
CC Name=1132;
CC Name=1133;
CC Name=1134;
CC Name=1135;
CC Name=1136;
CC Name=1137;
CC Name=1138;
CC Name=1139;
CC Name=1140;
CC Name=1141;
CC Name=1142;
CC Name=1143;
CC Name=1144;
CC Name=1145;
CC Name=1146;
CC Name=1147;
CC Name=1148;
CC Name=1149;
CC Name=1150;
CC Name=1151;
CC Name=1152;
CC Name=1153;
CC Name=1154;
CC Name=1155;
CC Name=1156;
CC Name=1157;
CC Name=1158;
CC Name=1159;
CC Name=1160;
CC Name=1161;
CC Name=1162;
CC Name=1163;
CC Name=1164;
CC Name=1165;
CC Name=1166;
CC Name=1167;
CC Name=1168;
CC Name=1169;
CC Name=1170;
CC Name=1171;
CC Name=1172;
CC Name=1173;
CC Name=1174;
CC Name=1175;
CC Name=1176;
CC Name=1177;
CC Name=1178;
CC Name=1179;
CC Name=1180;
CC Name=1181;
CC Name=1182;
CC Name=1183;
CC Name=1184;
CC Name=1185;
CC Name=1186;
CC Name=1187;
CC Name=1188;
CC Name=1189;
CC Name=1190;
CC Name=1191;
CC Name=1192;
CC Name=1193;
CC Name=1194;
CC Name=1195;
CC Name=1196;
CC Name=1197;
CC Name=1198;
CC Name=1199;
CC Name=1200;
CC Name=1201;
CC Name=1202;
CC Name=1203;
CC Name=1204;
CC Name=1205;
CC Name=1206;
CC Name=1207;
CC Name=1208;
CC Name=1209;
CC Name=1210;
CC Name=1211;
CC Name=1212;
CC Name=1213;
CC Name=1214;
CC Name=1215;
CC Name=1216;
CC Name=1217;
CC Name=1218;
CC Name=1219;
CC Name=1220;
CC Name=1221;
CC Name=1222;
CC Name=1223;
CC Name=1224;
CC Name=1225;
CC Name=1226;
CC Name=1227;
CC Name=1228;
CC Name=1229;
CC Name=1230;
CC Name=1231;
CC Name=1232;
CC Name=1233;
CC Name=1234;
CC Name=1235;
CC Name=1236;
CC Name=1237;
CC Name=1238;
CC Name=1239;
CC Name=1240;
CC Name=1241;
CC Name=1242;
CC Name=1243;
CC Name=1244;
CC Name=1245;
CC Name=1246;
CC Name=1247;
CC Name=1248;
CC Name=1249;
CC Name=1250;
CC Name=1251;
CC Name=1252;
CC Name=1253;
CC Name=1254;
CC Name=1255;
CC Name=1256;
CC Name=1257;
CC Name=1258;
CC Name=1259;
CC Name=1260;
CC Name=1261;
CC Name=1262;
CC Name=1263;
CC Name=1264;
CC Name=1265;
CC Name=1266;
CC Name=1267;
CC Name=1268;
CC Name=1269;
CC Name=1270;
CC Name=1271;
CC Name=1272;
CC Name=1273;
CC Name=1274;
CC Name=1275;
CC Name=1276;
CC Name=1277;
CC Name=1278;
CC Name=1279;
CC Name=1280;
CC Name=1281;
CC Name=1282;
CC Name=1283;
CC Name=1284;
CC Name=1285;
CC Name=1286;
CC Name=1287;
CC Name=1288;
CC Name=1289;
CC Name=1290;
CC Name=1291;
CC Name=1292;
CC Name=1293;
CC Name=1294;
CC Name=
```

DR GO; GO:0007219; P:N receptor signaling pathway; IC.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000152; Asx_hydroxyl.
 DR InterPro; IPR000742; EGF_2.
 DR InterPro; IPR001881; EGF_CA.
 DR InterPro; IPR001438; EGF_II.
 DR InterPro; IPR006209; EGF_Like.
 DR InterPro; IPR02049; Laminin_EGF.
 DR InterPro; IPR00800; Notch.
 DR Pfam; PF00023; ank; 6.
 DR Pfam; PF00008; EGF; 34.
 DR Pfam; PF00066; notch; 2.
 DR PRINTS; PR00010; EGFBLDOD.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR PRINTS; PR01452; NOTCH.
 DR SMART; SM00248; ANK; 6.
 DR SMART; SM00179; EGF_CA; 23.
 DR SMART; SM00004; NL_3.
 DR PROSITE; PS50086; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 22.
 DR PROSITE; PS00022; EGF_1; 33.
 DR PROSITE; PS01186; EGF_2; 27.
 DR PROSITE; PS01187; EGF_CA; 22.
 DR Receptor; Transcription regulation; Activator; Differentiation;
 KW Developmental protein; Repeat; ANK repeat; EGF-like domain;
 KW Transmembrane; Glycoprotein; Signal; Phosphorylation;
 KW Alternative splicing.
 KM SIGNAL
 FT CHAIN 1 25 POTENTIAL.
 FT CHAIN 26 2470 NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 FT CHAIN 1666 2470 NOTCH EXTRACELLULAR TRUNCATION.
 FT CHAIN 1697 2470 NOTCH EXTRACELLULAR DOMAIN.
 FT TRANSMEM 26 1677 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 1698 1698 POTENTIAL.
 FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 26 63 EGF-Like 1.
 FT DOMAIN 105 102 EGF-Like 2.
 FT DOMAIN 144 143 EGF-Like 3.
 FT DOMAIN 182 219 EGF-Like 4.
 FT DOMAIN 221 256 EGF-Like 5, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 258 294 EGF-Like 6 (INCOMPLETE).
 FT DOMAIN 296 334 EGF-Like 7, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 336 372 EGF-Like 8, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 373 411 EGF-Like 9, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 413 452 EGF-Like 10.
 FT DOMAIN 454 490 EGF-Like 11, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 492 528 EGF-Like 12, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 530 566 EGF-Like 13, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 568 603 EGF-Like 14, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 605 641 EGF-Like 15, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 643 678 EGF-Like 16, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 680 716 EGF-Like 17, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 718 753 EGF-Like 18, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 753 791 EGF-Like 19.
 FT DOMAIN 793 829 EGF-Like 20, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 831 869 EGF-Like 21, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 871 907 EGF-Like 22.
 FT DOMAIN 909 945 EGF-Like 23, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 947 983 EGF-Like 24, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 985 1021 EGF-Like 25, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1023 1059 EGF-Like 26, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1061 1097 EGF-Like 27, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1099 1145 EGF-Like 28.
 FT DOMAIN 1147 1183 EGF-Like 29.
 FT DOMAIN 1185 1221 EGF-Like 30, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1223 1260 EGF-Like 31, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1262 1300 EGF-Like 32, CALCITUM-BINDING (POTENTIAL).
 FT DOMAIN 1302 1345 EGF-Like 33.
 FT DOMAIN 1372 1410 EGF-Like 34.
 FT DOMAIN 1410 1454 EGF-Like 35.
 FT REPEAT 1501 1533 LIN/NOTCH 1.
 FT REPEAT 1825 1869 LIN/NOTCH 2.
 FT ANK 1.

FT REPEAT 1874 1903 ANK 2.
 FT REPEAT 1907 1937 ANK 3.
 FT REPEAT 1941 1970 ANK 4.
 Query Match 3.7%; Score 8; DB 1; Length 2470;
 Best Local Similarity 100.0%; Pred. No.32;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 120 AAAAPAP 127
 Db 2183 AAAAPAP 2190
 RESULT 15
 ANPA PSEAM STANDARD; PRT; 82 AA.
 ID ANPA PSEAM
 AC P04002;
 DT 23-OCT-1986 (Rel. 02, Created)
 DT 23-OCT-1986 (Rel. 02, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antifreeze protein A/B precursor.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; OC Acanthomorpha; Acanthopterygii; Perciformes; Pleuronectiformes; OC Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
 OX NCBI_TaxID=8265;
 RN [1]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=82197490; PubMed=6952188;
 RA Davies P.L., Roach A.H., Hew C.-L.;
 RT "DNA sequence coding for an antifreeze protein precursor from winter flounder."
 RL Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
 RN [2]
 RP SEQUENCE FROM N.A. (PROTEIN A).
 RX MEDLINE=88259236; PubMed=313486;
 RA Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
 RT "Differential amplification of antifreeze protein genes in the Pleuronectinae."
 RL J. Mol. Evol. 27:29-35(1988).
 RN [3]
 RP SEQUENCE FROM N.A. (PROTEIN B).
 RX MEDLINE=84264559; PubMed=6086629;
 RA Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
 RT "Antifreeze protein genes of the winter flounder."
 RL J. Biol. Chem. 259:9241-9247(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92209995; PubMed=1555765;
 RA Davies P.L.;
 RT "Conservation of antifreeze protein-encoding genes in tandem repeats."
 RL Gene 112:163-170(1992).
 RN [5]
 RP 3D-STRUCTURE MODELING OF 45-81.
 RX MEDLINE=92148833; PubMed=1738160;
 RA Chou K.-C.;
 RT "Energy-optimized structure of antifreeze protein and its binding mechanism."
 RL J. Mol. Biol. 223:509-517(1992).
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
 RX MEDLINE=95281060; PubMed=7760940;
 RA Sichert P., Yang D.S.C.;
 RT "Ice-binding structure and mechanism of an antifreeze protein from winter flounder."
 RL Nature 375:427-431(1995).
 CC -1- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
 CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF PROTEIN A.
 CC -1- SIMILARITY: BELONGS TO THE TYPE-I AFP FAMILY. TYPE 1 AFP ARE ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L00138; AAB59964.1; -;
DR EMBL; L29178; AAB59964.1; JOINED.
DR EMBL; M62414; AAA49469.1; -;
DR EMBL; X07506; CAA30389.1; -;
DR EMBL; M62416; AAA49471.1; -;
DR EMBL; M62417; AAA49472.1; -;
DR PIR; A05161; A05161.
DR PIR; J50704; PDPLAN.
DR PIR; J50706; J50706.
DR PIR; S02326; S02326.
DR PDB; 1ATF; 15-OCT-94.
DR PDB; 1WFA; 03-JUN-95.
DR PDB; 1WFB; 03-JUN-95.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT PROPEP 22 44
FT CHAIN 45 82 REMOVED BY A DIPEPTIDYLPEPTIDASE
FT VARIANT 36 36 (PROBABLY).
FT VARIANT 70 70 ANTIFREEZE PROTEIN A/B.
FT CONFLICT 24 24 A -> V.
FT HELIX 46 80 A -> D (IN PROTEIN B).
SQ SEQUENCE 82 AA; 7711 MW; C2AB7B74C0D46CC1 CRC64;
S -> R (IN REF. 2).

Query Match 3.28; Score 7; DB 1; Length 82;
Best Local Similarity 100.0%; Pred.No.15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAA 126
DB 35 AAAAPAA 41

Search completed: January 15, 2004, 15:29:01
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:26:21 ; Search time 35 Seconds
(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MGLLMACIVCAFAFRKRRF.....TAKEPAEPHPSPEIQANQ 219

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 830525 seqs, 258052604 residues

Word size: 0

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	9	4.1	73	5	Q8MSH9	Q8MSH9 drosophila
2	9	4.1	75	5	Q9VB16	Q9VB16 drosophila
3	9	4.1	344	11	Q9DSJ3	Q9DSJ3 mus musculus
4	9	4.1	784	16	Q8PH14	Q8PH14 xanthomonas
5	9	4.1	841	5	Q9NEA5	Q9NEA5 leishmania
6	9	4.1	2969	16	Q8NS46	Q8NS46 corynebacter
7	8	3.7	125	10	Q8H4F8	Q8H4F8 oryza sativa
8	8	3.7	139	16	Q8Y2D2	Q8Y2D2 raietonia b
9	8	3.7	141	16	Q8Y2R3	Q8Y2R3 raietonia b
10	8	3.7	145	10	Q9FPC0	Q9FPC0 arabidopsis
11	8	3.7	145	10	Q9F516	Q9F516 arabidopsis
12	8	3.7	193	10	Q9C7F7	Q9C7F7 arabidopsis
13	8	3.7	193	10	Q8LE10	Q8LE10 arabidopsis
14	8	3.7	209	15	Q9DCU8	Q9DCU8 human immun
15	8	3.7	228	16	Q9AKA2	Q9AKA2 caulobacter
16	8	3.7	231	2	Q93ND2	Q93ND2 myxococcus

17	8	3.7	233	2 Q8GCE5	Q8GCE5 streptomyces
18	8	3.7	242	4 Q8NRC8	Q8NRC8 homo sapien
19	8	3.7	259	2 Q91UC4	Q91UC4 rhizobium m
20	8	3.7	262	10 Q8W2Q2	Q8W2Q2 oryza sativ
21	8	3.7	266	16 Q8P710	Q8P710 xanthomonas
22	8	3.7	268	16 P95192	P95192 mycobacteri
23	8	3.7	272	10 Q8S2S9	Q8S2S9 thelningiel
24	8	3.7	276	16 Q92WH2	Q92WH2 rhizobium m
25	8	3.7	285	11 Q8BHC2	Q8BHC2 mus musculu
26	8	3.7	291	10 Q40478	Q40478 nicotiana t
27	8	3.7	291	10 Q8H382	Q8H382 oryza sativ
28	8	3.7	302	5 Q9N9A5	Q9N9A5 leishmania
29	8	3.7	325	2 Q922X8	Q922X8 fraetulia s
30	8	3.7	336	16 Q9AB24	Q9AB24 caulobacter
31	8	3.7	352	11 Q9CX00	Q9CX00 mus musculu
32	8	3.7	364	16 Q8B885	Q8B885 xanthomonas
33	8	3.7	380	11 Q8B2V6	Q8B2V6 mus musculu
34	8	3.7	381	2 Q86996	Q86996 acinetobact
35	8	3.7	384	16 Q9AD02	Q9AD02 streptomyces
36	8	3.7	401	8 Q8ME27	Q8ME27 primula jes
37	8	3.7	428	10 Q94108	Q94108 zea mays fm
38	8	3.7	434	10 Q8H385	Q8H385 oryza sativ
39	8	3.7	443	8 Q8ME20	Q8ME20 primula pol
40	8	3.7	448	8 Q31310	Q31310 streptitza
41	8	3.7	468	8 Q9GEC0	Q9GEC0 primula cor
42	8	3.7	468	8 Q8ME93	Q8ME93 primula bul
43	8	3.7	493	2 Q48431	Q48431 klebsiella
44	8	3.7	493	5 Q9KAP4	Q9KAP4 drosophila
45	8	3.7	493	5 Q8S2K1	Q8S2K1 drosophila

ALIGNMENTS

RESULT 1

ID	Q8MSH9	PRELIMINARY;	PRT;	73 AA.
AC	Q8MSH9;			
DT	01-OCT-2002 (T-REMBLrel. 22, Created)			
DT	01-OCT-2002 (T-REMBLrel. 22, Last sequence update)			
DT	01-MAR-2003 (T-REMBLrel. 23, Last annotation update)			
DE	GH19893P (Fragment).			
GN	BCDNA:GH19893.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=Berkely;			
RA	Strapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,			
RA	Champe M., Chavez C., Dorsett V., Dreshnek D., Farfan D., Frise E.,			
RA	George R., Gonzalez M., Guartin H., Kronmiller B., Li P., Liao G.,			
RA	Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,			
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,			
RA	Celniker S.;			
RU	Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AY118602; AAM50662.1; -			
DR	FLYBase; FBgn0063235; BCDNA:GH19893.			
FT	NON TER			
SO	SEQUENCE. 73 AA; 6847 MW; BC8B7C976694F2D CRC64;			

Query Match 4.1%; Score 9; DB 5; Length 73;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAAPAAP 128
Db 43 AAAAPAAP 51

RESULT 2

09YBL6 PRELIMINARY; PRT; 75 AA.
 ID 09YBL6 Q24390;
 AC 09YBL6 Q24390 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Accessory gland-specific peptide 57Da precursor (Male accessory gland secretory protein 57Da).
 GN M57DA OR CG9074.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OC NCB1_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
 RC STRAIN=OREGON-R;
 RX MEDLINE=95227188; PubMed=7711745;
 RA Stimmerl E., Schaefer M., Schaefer U.;
 RA "Structure and regulation of a gene cluster for male accessory gland transcript in Drosophila melanogaster";
 RT Insect Biochem. Mol. Biol. 25:127-137(1995).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196066; PubMed=10731132;
 RA Adams M.D., Cealiker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Flamkoc C., Baldwin D., Bailew R.M., Baeu A., Bakendal J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borikava D., Botchan M.R., Butler H., Brokstein P., Brotier P., Burts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K., Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C., Jafarzadeh S., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Miletina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M., Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T., Spier B., Spradling A.C., Stapleton M., Strong R., Sun E., Switzkae R., Tector C., Turner R., Venter B., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach U., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of Drosophila melanogaster";
 RT Science 287:2185-2195(2000).
 RL [1]
 RP FUNCTION: THIS PROTEIN IS TRANSFERRED FROM MALE TO FEMALE DURING MATING AND MAY AFFECT EGG-LAYING AND BEHAVIOR AFTER MATING.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- TISSUE SPECIFICITY: LUMEN FLUID OF MALE ACCESSORY GLANDS, BECOMES SEMINAL FLUID.
 CC EMBL: Z33647; CAA83925.1; -;
 DR EMBL: AE003753; AAF56515.1; -;
 DR FlyBase: FBgn0011668; Maf57Da.
 KW Signal; Behavior.

FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 75 ACCESSORY GLAND-SPECIFIC PEPTIDE 57DA.
 FT DOMAIN 39 64 ALA/PRO-RICH DOMAIN.
 FT CONFLICT 39 46 MISSING (IN REF. 1).
 FT CONFLICT 64 75 MISSING (IN REF. 1).
 SQ SEQUENCE 75 AA; 7106 MW; BCD3D70817C98E14 CRC64;
 Query Match 4.1%; Score 9; DB 11; Length 75;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 120 AAAAPAPP 128
 DB 45 AAAAPAPP 53
 RESULT 3
 ID 09D5J3 PRELIMINARY; PRT; 344 AA.
 AC 09D5J3
 DT 01-JUN-2001 (TRENBLrel. 17, Created)
 DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
 DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
 DE 4930432K03R1X protein.
 GN 4930432K03R1X.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Testis;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawachi T., Shimagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Akawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T., Fleischmann W., Gasteirland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo Y., Nkaido I., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M., Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombere P., Norone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitteker C., Wilming L., Wymahar-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohlsunki S., Hayashizaki Y.;
 RA "Functional annotation of a full-length mouse cDNA collection";
 RT Nature 409:685-690(2001).
 RL EMBL: AK015291; BAB29782.1; -;
 DR MGI: MGI:1921029; 4930432K03R1X.
 DR InterPro: IPR002965; P rich exten.
 DR PRINTS: PR01217; PRICHTEN.
 SQ SEQUENCE 344 AA; 36294 MW; D65EAD71CE5802AC CRC64;
 Query Match 4.1%; Score 9; DB 11; Length 344;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLIMACT 9
 DB 1 MGLIMACT 9
 RESULT 4
 ID 08PH14 PRELIMINARY; PRT; 784 AA.
 AC 08PH14
 DT 01-OCT-2002 (TRENBLrel. 22, Created)

```

DT 01-OCT-2002 (TEMBLrel. 22, last sequence update)
DE Hypothetical protein XAC3446.
GN XAC3446.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.B.A.,
RA Camarotte G., Camargo F., Cardozo J., Chambeiro F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Subbal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AE011991; AM38289.1; -.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00486; trans_reg_Cf_1.
DR Prodom; PD000329; Trans_reg_Cf_1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 784 AA; 84093 MW; F41377005382EC8 CRC64;

QY 152 APVAAPPA 160
Db 160 APVAAPPA 168

RESULT 5
QNEAS PRELIMINARY; PRT; 841 AA.
AC QNEAS;
DT 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE Possible inner dynein arm.
GN L5515.03.
OS Leishmania major.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OC NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Masny D., Purnelle B., Goffeau A., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RX MEDLINE=9816435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
DR Genome Res. 8:135-145(1998).
EMBL; AL161414; CAB7677.1; -.

```

```

DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 4.
DR SMART; SM00320; WD40; 4.
DR PROSITE; PS00678; WD_REPEATS_1; 2.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
SQ SEQUENCE 841 AA; 93125 MW; B41B72F1BB821226 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 5; Length 841;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAP 127
Db 40 SAAAPAP 48

RESULT 6
QNS46 PRELIMINARY; PRT; 2969 AA.
AC QNS46;
DT 01-OCT-2002 (TEMBLrel. 22, Created)
DT 01-OCT-2002 (TEMBLrel. 22, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE 3-oxoacyl-(acyl-carrier-protein) synthase (EC 2.3.1.85).
GN CGL0836.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005276; BAB98229.1; -.
DR InterPro; IPR001227; Ac transferase.
DR InterPro; IPR003965; Fatty acid synth.
DR InterPro; IPR000794; Ketoacyl-synt.
DR InterPro; IPR002539; Maoc dehydratas.
DR InterPro; IPR002016; Peroxidase.
DR Pfam; PF00698; Acyl_transf_1.
DR Pfam; PF00109; ketoacyl-synt_1.
DR Pfam; PF02801; ketoacyl-synt_1.
DR Pfam; PF01575; Maoc dehydratas; 1.
DR PRINTS; PRO1483; PASTINASE.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE; PS00435; PEROXIDASE_1; 1.
DR TRANSFERASE; Acyltransferase; Complete proteome.
SQ SEQUENCE 2969 AA; 315124 MW; 908B45A60D90815 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 2969;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAPAPVA 176
Db 1663 PAAPAPVA 1671

RESULT 7
QNS46 PRELIMINARY; PRT; 125 AA.
AC QNS46;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, last annotation update)
DE Cj1221 H04.16 protein.
GN Cj1221 H04.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;

```

```

OC Ehirarioidae; Oryzae; Oryza.
OX NCB1_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.,
RT "Oryza sativa nipponbare (GAJ) genomic DNA, chromosome 8, BAC
  clone:O1221.H04."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF004183; BAC21432.1; -.
SQ SEQUENCE 125 AA; 13560 MW; 31F75D8398B8A52 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 15;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAP 128
DB 11 AAAPAAP 18

RESULT 8
Q8Y2D2 PRELIMINARY; PRT; 139 AA.
AC Q8Y2D2;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical signal peptide protein RSC0404.
GN RSC0404 OR RS03372.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Ralstoniaceae; Ralstonia.
RN NCB1_TaxID=305;
RP [1]
RC SEQUENCE FROM N.A.
RX STRAIN=GM11000;
MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
  Arlet M., Billault A., Brottier P., Camus J.C., Catrolicco L.,
  Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
  Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
  Signier P., Thebaud P., Whalen M., Winkler P., Levy M.,
  Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AF646059; CAD13932.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 139 AA; 14153 MW; A3150B6FD629039 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127
DB 55 AAAPAAP 62

RESULT 9
Q8Y2R3 PRELIMINARY; PRT; 141 AA.
AC Q8Y2R3;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein RSC0272.
GN RSC0272 OR RS03239.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OX Ralstoniaceae; Ralstonia.
RN NCB1_TaxID=305;
RP [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
  Arlet M., Billault A., Brottier P., Camus J.C., Catrolicco L.,
  Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
  Gaspin C., Lave M., Moisan A., Robert C., Saurin W., Schlex T.,
  Signier P., Thebaud P., Whalen M., Winkler P., Levy M.,
  Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL; AF646058; CAD13800.1; -.
DR InterPro; IPR001763; Rhodanese-like.
DR SMART; SM00450; RHOD; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 15454 MW; 878FE1840A8B84E CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156
DB 126 AAGAPVAA 133

RESULT 10
Q9PFC0 PRELIMINARY; PRT; 145 AA.
AC Q9PFC0;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Histone H2B like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCB1_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Columbia;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
  Miyajima N., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
  features of the 1.6 Mb regions covered by twenty physically assigned
  BAC clones."
RL DNA Res. 4:215-230(1997).
CC -1 SUBUNIT: THE NUCLEOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
  H2A, H2B, H3, AND H4, WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
  SIMILARITY).
CC -1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1 SIMILARITY: BELONGS TO THE HISTONE H2B FAMILY.
DR EMBL; AB005243; BAB10609.1; -.
DR InterPro; IPR004822; Histone core.
DR InterPro; IPR000558; Histone_H2B.
DR Pfam; PF00125; histone_1.
DR PRINTS; PR00621; HISTONEH2B.
DR ProDom; PD000497; Histone_H2B; 1.
DR SMART; SM00427; H2B; 1.
DR PROSITE; PS00357; HISTONE_H2B; 1.
KW Chromosomal protein; DNA-Binding; Nuclear protein; Nucleosome core.
SQ SEQUENCE 145 AA; 15732 MW; CC6421B559D42B30 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AABPAAA 137
DB 19 AABPAAA 26

```

RESULT 11
 096516 PRELIMINARY; PRT; 145 AA.
 AC 096516;
 DT 01-FEB-1997 (TREMBlrel. 02, Created)
 DT 01-FEB-1997 (TREMBlrel. 02, last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, last annotation update)
 DE Histone H2B like protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ecotype columbia;
 RA Phillips G.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: THE NUCLEOSOME IS AN OCTAMER CONTAINING TWO MOLECULES OF
 CC H2A, H2B, H3, AND H4; WHICH WRAP APPROXIMATELY 146 BP OF DNA (BY
 CC SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
 CC EMBL; Y07745; CAA69025.1; -
 DR InterPro; IPR004822; Histone core.
 DR InterPro; IPR000558; Histone_H2B.
 DR Pfam; PF00125; histone.1.
 DR PRINTS; PRO0621; HISTONEH2B.
 DR PRODOM; PD000497; Histone_H2B; 1.
 DR SMART; SM00427; H2B; 1.
 DR PROSITE; PS00357; HISTONE_H2B; 1.
 KM Chromosomal protein; DNA-binding; Nuclear protein; Nucleosome core.
 SQ SEQUENCE 145 AA; 1573 MW; 6C8EB818390F4666 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 145;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 130 AAAPAAP 137
 Db 19 AAAPAAP 26

RESULT 12
 09677 PRELIMINARY; PRT; 193 AA.
 AC 09677;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
 DE Lipid transfer protein, putative.
 GN F13K9.6 OR AT1G37950.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA MEDLINE=21016719; PubMed=11130712;
 RA Theologis A.; Ecker J.R.; Palm C.J.; Federapfel N.A.; Kaul S.;
 RA White O.; Alonso J.; Altafaj H.; Araujo R.; Bowman C.L.; Brooks S.Y.;
 RA Buehler E.; Chan A.; Chao Q.; Chen H.; Cheuk R.F.; Chin C.W.;
 RA Chung M.K.; Conn L.; Conway A.B.; Creasy T.H.; Dewar K.;
 RA Dunn P.; Egu P.; Feldblum T.V.; Feng J.-D.; Fong B.; Fujii C.Y.;
 RA Gali J.E.; Goldsmith A.D.; Haas B.; Hansen N.F.; Hughes B.; Hultzar L.;
 RA Hunter J.L.; Jenkins J.; Johnson-Hopson C.; Khan S.; Khaykin E.;
 RA Kim C.J.; Koo H.L.; Kremenetskaia I.; Kurtz D.B.; Kwan A.; Lam B.;
 RA Langin-Hooper S.; Lee A.; Lee J.M.; Lenz C.A.; Li J.H.; Li Y.-P.;
 RA Lin X.; Liu S.X.; Liu Z.A.; Luros J.S.; Malti R.; Marziani A.;
 RA Miltscher J.; Miranda M.; Nguyen M.; Nieman W.C.; Osborne B.T.;

PAI G.; Peterson J.; Pham P.K.; Rizzo M.; Rooney T.; Rowley D.;
 RA Sakano H.; Salzer S.L.; Schwartz J.R.; Shinn P.; Southwick A.M.;
 RA Sun R.; Tallon L.J.; Tambunga G.; Toriumi M.J.; Town C.D.;
 RA Utecherack T.; Van Aken S.; Vaynsberg M.; Vysotskaya V.S.; Walker M.;
 RA Wu D.; Yu G.; Fraser C.M.; Venter J.C.; Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 RT thaliana.";
 RL Nature 408:816-820(2000).
 RN [2]

RP SEQUENCE FROM N.A.
 RA Southwick A.; Karlin-Neumann G.; Nguyen M.; Lam B.; Miranda M.;
 RA Palm C.J.; Bower L.; Jones T.; Banh J.; Carninci P.; Chen H.;
 RA Cheuk R.; Chung M.K.; Hayashizaki Y.; Ishida J.; Kamuya A.; Kawai J.;
 RA Kim C.; Lin J.; Liu S.X.; Narusaka M.; Pham P.K.; Sakano H.;
 RA Sakurai T.; Satou M.; Seki M.; Shinn P.; Yamada K.; Shinozaki K.;
 RA Ecker J.; Theologis A.; Davis R.W.;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
 RA Tripp M.; Southwick A.; Karlin-Neumann G.; Nguyen M.; Miranda M.;
 RA Palm C.J.; Bower L.; Jones T.; Banh J.; Carninci P.; Chen H.;
 RA Cheuk R.; Chung M.K.; Hayashizaki Y.; Ishida J.; Kamuya A.; Kawai J.;
 RA Kim C.; Lin J.; Liu S.X.; Narusaka M.; Pham P.K.; Sakano H.;
 RA Sakurai T.; Satou M.; Seki M.; Shinn P.; Yamada K.; Shinozaki K.;
 RA Ecker J.; Theologis A.; Davis R.W.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 RA EMBL; AC069471; AAG51485.1; -
 DR EMBL; AY092956; AAM12955.1; -
 DR EMBL; AY128712; AAM91112.1; -
 DR InterPro; IPR003612; AAI.
 DR Pfam; PF00234; tryd_alpha_amy1; 1.
 DR EMBL; AY128712; AAM91112.1; -
 DR InterPro; IPR003612; AAI.
 DR Pfam; PF00234; tryd_alpha_amy1; 1.
 SQ SEQUENCE 193 AA; 19759 MW; D54B38B12FEE6610 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 193;
 Best Local Similarity 100.0%; Pred. No. 22;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 120 AAAPAAP 127
 Db 20 AAAPAAP 27

RESULT 13
 09LE10 PRELIMINARY; PRT; 193 AA.
 AC 09LE10;
 DT 01-OCT-2002 (TREMBlrel. 22, Created)
 DT 01-OCT-2002 (TREMBlrel. 22, last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, last annotation update)
 DE Lipid transfer protein, putative.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Haas B.J.; Volfovsky N.; Town C.D.; Troukhan M.; Alexandrov N.;
 RA Feldmann K.A.; Flavell R.B.; White O.; Salzberg S.L.;
 RT "Full-length messenger RNA sequences greatly improve genome
 RT annotation.";
 RL Genome Biol. 0:0-0(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brover V.; Troukhan M.; Alexandrov N.; Lu Y.-P.; Flavell R.;
 RA Feldmann K.;
 RT "Full-length cDNA from Arabidopsis thaliana.";
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY085407; AAM62634.1; -
 DR InterPro; IPR003612; AAI.
 DR Pfam; PF00234; tryd_alpha_amy1; 1.
 DR SMART; SM00499; AAI; 1.
 SQ SEQUENCE 193 AA; 19747 MW; C091074205D76610 CRC64;

Query Match 3.7%; Score 8; DB 10; Length 193;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 20 AAAAPAP 27

RESULT 14

Q9D0U8 PRELIMINARY; PRT; 209 AA.
AC Q9D0U8;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE Nef protein (Negative factor) (27 kDa protein).
GN NEF.
OS Human immunodeficiency virus 1.
OC Viruses; Retrod viruses; Retroviridae; Lentivirus.
NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=01u09;
RX MEDLINE=21002575; PubMed=11118071;
RA Gaffin R., Wolf D., Miller R., Hill M.D., Stellwag B., Freitag M.,
Sas G., Scott G.B., Baur A.S.;
RT "Functional and structural defects in HIV-1 nef genes derived from
pediatric long-term survivors."
RL AIDS Res. Hum. Retroviruses 16:1855-1868(2000).
CC -!- FUNCTION: NEF HAS GTPASE, GTP-BINDING AND AUTOPHOSPHORYLATING
ACTIVITIES. IT SEEMS TO DOWN-REGULATE THE CD4 (T4) ANTIGEN (BY
SIMILARITY).
DR EMBL; AF252901; AAC34573.1; -.
DR HSP; P03406; IEFN.
DR InterPro; IPR001558; HIV_Nef.
DR Pfam; PF00469; F-protein; 1.
DR ProDom; PD000031; HIV_Nef; 1.
KM AIDS; GTP-binding; Lipoprotein; Myristate.
SQ SEQUENCE 209 AA; 23656 MW; 6B858BAD98C79519 CRC64;

Query Match 3.7%; Score 8; DB 15; Length 209;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 131 ABPAPAP 138
DB 23 ABPAPAP 30

RESULT 15

Q9A4A2 PRELIMINARY; PRT; 228 AA.
AC Q9A4A2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Cytochrome c, membrane-bound.
GN CC2935.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.B.,
Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
Koloney J.F., Smit J., Craven M.B., Knouri H., Shetty J., Berry K.,

RA Uterback T., Tran K., Wolf A., Vamathavan J., Ermolaeva M., White O.,
RA Salizberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AR005958; AK24897.1; -.
DR HSP; P81458; I154.
DR TIGR; CC2935; -.
DR InterPro; IPR001014; Antifreeze_1.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR003088; Cyt_C1.
DR InterPro; IPR002327; Cyt_C1AB.
DR Pfam; PF000034; cytochrome_c; 1.
DR PRINTS; PR00308; ANTI-FREEZE1.
DR PRINTS; PR00604; CYTCHROME_C1AB.
DR ProDom; PD000375; Cyt_C1AB; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 1.
KM Complete proteome.
SQ SEQUENCE 228 AA; 23023 MW; 5AB05C7DD5D05F82 CRC64;

Query Match 3.7%; Score 8; DB 16; Length 228;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
DB 219 AAAAPAP 226

Search completed: January 15, 2004, 15:29:49
Job time : 35 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:26:56 ; Search time 21 Seconds
(without alignments)
1002.902 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219

Sequence: 1 MKLLMAGIVCAVAFARRRRF.....TAKPADEPHPSLEQANQ 219

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 283308 seqs, 96168682 residues

Word size: 0

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: 1: PIR 76:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.1	581	2	probable glutamyl-nitrate reductase
2	9	4.1	864	1	JC1422
3	8	3.7	110	1	R6BYP3
4	8	3.7	129	2	H86500
5	8	3.7	129	2	C72122
6	8	3.7	193	2	H86404
7	8	3.7	228	2	H87612
8	8	3.7	268	2	G70645
9	8	3.7	276	2	G95887
10	8	3.7	291	2	T02434
11	8	3.7	325	2	T48873
12	8	3.7	336	2	A67300
13	8	3.7	351	2	T46827
14	8	3.7	383	2	T46827
15	8	3.7	396	1	JH0633
16	8	3.7	495	2	T52066
17	8	3.7	506	2	A40679
18	8	3.7	523	2	B40679
19	8	3.7	550	2	C75557
20	8	3.7	574	2	S23530
21	8	3.7	611	2	C96030
22	8	3.7	846	2	S52418
23	8	3.7	1476	2	A45773
24	8	3.7	4307	2	T20721
25	8	3.2	78	2	S33173
26	8	3.2	78	2	B83446
27	7	3.2	82	1	FDPLAW
28	7	3.2	82	2	S02326
29	7	3.2	82	2	A05161

30	7	3.2	82	2	T51125	antifreeze protein
31	7	3.2	100	2	F95852	hypothetical prote
32	7	3.2	101	2	H69091	ribosomal protein
33	7	3.2	102	2	D64363	ribosomal protein
34	7	3.2	106	1	R6DOP2	acidic ribosomal p
35	7	3.2	107	1	R6KMLC	acidic ribosomal p
36	7	3.2	109	1	R6JUMP1	acidic ribosomal p
37	7	3.2	110	2	T37490	ribosomal protein
38	7	3.2	111	2	E72524	probable ribosomal
39	7	3.2	112	1	B95857	hypothetical prote
40	7	3.2	112	2	R5FP28	acidic ribosomal p
41	7	3.2	112	2	S54179	acidic ribosomal p
42	7	3.2	112	2	E86141	protein T25K16.9
43	7	3.2	114	1	R5KTI1	acidic ribosomal p
44	7	3.2	114	1	R5HS2H	ribosomal protein
45	7	3.2	114	2	F84266	50S ribosomal prot

ALIGNMENTS

RESULT 1

probable glutamyl-tRNA reductase - Streptomyces coelicolor

C/Species: Streptomyces coelicolor

C/Date: 03-Dec-1999 #Sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T36267

R/Murphy, L.; Harries, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A/Reference number: Z21576

A/Accession: T36267

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-581 <MUR>

A/Cross-references: EMBL:AL079345; PIDN:CA845353.1; GSPDB:GN00070; SCOPDB:SC68.17C

A/Experimental source: strain A3(2)

C/Genetics:

A/Genes: SCOPDB:SC68.17C

Query Match

Best Local Similarity 100.0%; Pred. No. 2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 154 VAAAPAAEA 162

Db 36 VAAAPAAEA 44

RESULT 2

JC1422

nitrate reductase (NADH) (EC 1.7.1.1) - Volvox carter1

C/Species: Volvox carter1

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Jun-2002

C/Accession: JC1422; S22192

R/Gruber, H.; Goeltz, S.D.; Kirk, D.L.; Schmitt, R.

Gene 120, 75-83, 1992

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

antifreeze protein
hypothetical prote
ribosomal protein
ribosomal protein
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
acidic ribosomal p
ribosomal protein
probable ribosomal
hypothetical prote
acidic ribosomal p
acidic ribosomal p
protein T25K16.9
acidic ribosomal p
ribosomal protein
50S ribosomal prot

F:613-864/Domain: cytochrome-b5 reductase homology

F:139/Binding site: molibdopterin (Cys) (covalent) #status predicted
F:532/Binding site: heme iron (His) (axial ligands) #status predicted

Query Match 4.1%; Score 9; DB 1; Length 864;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 128
DB 478 AAAAPAAP 486

RESULT 3

606 acidic ribosomal protein p1-alpha - fission yeast (*Schizosaccharomyces pombe*)
N/Alternate names: acidic ribosomal protein 3; ribosomal protein Y12a11
C/Species: *Schizosaccharomyces pombe*
C/Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 10-Dec-1999
C/Accession: C34715; T40352
R/Beltrame, M.; Bianchi, M.E.
Mol. Cell. Biol. 10, 2341-2348, 1990
A>Title: A gene family for acidic ribosomal proteins in *Schizosaccharomyces pombe*: two
A/Reference number: A34715; MUID:90220620; PMID:2325655
A/Accession: C34715
A/Molecule type: DNA
A/Residues: 1-110
A/Cross-references: EMBL:MA3139; NID:G173467; PIDN:AAA3536.1; PID:G173468
R/Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, March 1997
A/Reference number: Z21922
A/Accession: T40352
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-110 <NOO>
A/Cross-references: EMBL:AL022070; PIDN:CAA17793.1; GSPDB:GN00067; SPDB:SPBC39.13C
A/Experimental source: strain 972h-; coamid c389
C/Genetics:

A/Gene: rpa3
A/Map position: 2
A/Introns: 23/3; 103/3
C/Suprafamily: rat acidic ribosomal protein p1
C/Keyword: phosphoprotein; protein biosynthesis; ribosome

Query Match 3.7%; Score 8; DB 1; Length 110;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAAAPAA 126
DB 64 SAAAAAPAA 71

RESULT 4

L7/L12 ribosomal protein [imported] - *Chlamydomonas reinhardtii* (strain J138)
C/Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: H86500
R/Shirai, M.; Hatakeyama, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A>Title: Comparison of whole genome sequences of *Chlamydomonas reinhardtii* J138.
A/Reference number: A86491; MUID:2030349; PMID:10871362
A/Accession: H86500
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-129 <STO>
A/Cross-references: GB:BA000008; NID:G8978453; PIDN:BA98290.1; GSPDB:GN00142
A/Experimental source: strain J138
C/Genetics:

A/Gene: r17
C/Suprafamily: *Bacterichia coli* ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 BAPVAAAP 178
DB 50 BAPVAAAP 57

RESULT 5

ribosomal protein L7/L12 CP0695 [imported] - *Chlamydomonas reinhardtii* (strain CWL029)
C/Species: *Chlamydomonas reinhardtii*, *Chlamydomonas reinhardtii*
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C/Accession: C72122; G81548
R/Kalman, S.; Mitchell, W.; Marathe, R.; Lamell, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A>Title: Comparative genomes of *Chlamydomonas reinhardtii* and *C. trichomatis*.
A/Reference number: A72000; MUID:99206606; PMID:10192388
A/Accession: C72122
A/Molecule type: DNA
A/Residues: 1-129 <ARN>
A/Cross-references: GB:AE001593; GB:AE001363; NID:G4376334; PIDN:AA018233.1; PID:G43763
A/Experimental source: strain CWL029
R/Reed, T.D.; Brumham, R.C.; Shen, C.; Gyll, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg
Nucleic Acids Res. 28, 1397-1406, 2000
A>Title: Genome sequences of *Chlamydomonas trichomatis* MOPN and *Chlamydomonas reinhardtii* AR39.
A/Reference number: AB1500; MUID:20150255; PMID:10664935
A/Accession: G81548
A/Molecule type: DNA
A/Residues: 1-129 <REA>
A/Cross-references: GB:AE002228; GB:AE002161; NID:G7189606; PIDN:AA038503.1; PID:G71896
A/Experimental source: strain AR39, HL cells
C/Genetics:

A/Gene: r17; CP0695
C/Suprafamily: *Bacterichia coli* ribosomal protein L12

Query Match 3.7%; Score 8; DB 2; Length 129;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 BAPVAAAP 178
DB 50 BAPVAAAP 57

RESULT 6

probable lipid transfer protein [imported] - *Arabidopsis thaliana*
C/Species: *Arabidopsis thaliana* (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C/Accession: H86404
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federgräb, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marshall
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A/Reference number: AB6141; MUID:21016719; PMID:11130712
A/Accession: H86404
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-193 <STO>
A/Cross-references: GB:AE005172; NID:G11024871; PIDN:AA026955.1; GSPDB:GN00141
C/Genetics:

A/Map position: 1
Query Match 3.7%; Score 8; DB 2; Length 193;

Best Local Similarity 100.0%; Pred. NO. 6.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 20 AAAPAP 27

RESULT 7
E87612

C:\Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 18-Nov-2002
C:\Accession: E87612

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249, MUID:21173698, PMID:11259647

A:Residues: 1-228 <STO>
 A:Cross-references: GB:AE005673, NID:g13424561, PIDN:AAK24897.1, GSPDB:GN00148
 C:Genetics:

F:85/Binding site: heme iron (His) (axial ligand) #status predicted
F:150/Binding site: heme iron (Met) (axial ligand) #status predicted
C:Superfamily: membrane-bound cytochrome c/cy1; cytochrome c homolog
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:81.84/Binding site: heme (Cys) (covalent) #status predicted
F:85/Binding site: heme iron (His) (axial ligand) #status predicted
F:150/Binding site: heme iron (Met) (axial ligand) #status predicted

Best Local Similarity	100.0%;	Pred. No. 7.9;
Matches	8;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	219	AAAAPAP	226
x7			

RESULT 8
G70645

[illegible]

A::Title: Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome
A::Reference number: A70500; MUID:98295987; PMID:9634230
A::Accession: G70645

A:Residues: 1-268 <COL>
A:Molecule type: DNA
A:Cross-references: GR.Z81857, GR.A1123456, NTD.G3261695, PDBM.CAB06280 1, DT

A:Gene: RV3134C
C:Genetics:
A:Experimental source: strain H3/RV

Query Match 3.7%; Score 8; DB 2; length 268
Best Local Similarity 100.0%; Pred. No. 9.1;

match	0	CONSTANT	0	INDICATOR	0	GAPS	0
by	193	PSPAPAT	200				

27	133		200
Db	138	PSPABPAT	145

RESULT 9
G95887

CiDate: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 01-Mar-2002
CiAccession: G95887

A1>Title: The complete sequence of the 1,663-kb pSymB megaplasmid from the N2-fixing endo
A,Reference number: A95042; MUID:21396508; PMID:11481431

A: Molecule type: DNA
A: Residues: 1-276 <KUR>

pela, D., Chaitin, P., Cowle, A., Davis, R.W., Dreano, S., Federspiel, N.A., Fisher, R., L., Hyman, R.W., Jones, T. *Science* 293, 668-672, 2001

A;Reference number: A96039; WCID:21368234; PMID:11474104
A;Contents: annotation
C;Genetics:

A:Genome: plamid
C:Superfamily: spermidine/putrescine transport system permease protein potI
Query Match 3.7%; Score 8; DB 2; Length 276;
Best Local Similarity 100.0%; Pred. No. 9.3;

QY 130 AAEPAAAA 137

RESULT 10

C:Species: *Nicotiana tabacum* (common tobacco)
 C:sequence revision 05-Mar-1999 #text change 31-Mar-2001
 C:Date: 05-Mar-1999
 C:binding protein SBBF-1 - common tobacco

Plant Cell, 1: 179-182, 1993
A/Title: Ethylene-inducible DNA binding proteins that interact with an ethylene response element
A/Reference number: 214671; MUID: 95276459; PMID: 7756828
A/Accession: T02434

A. Cross-references: EMBL:D38125, NID:g7930361, PIDN:BAA07323.1, PID:G1208457
A. Experimental source: strain BY4, tissue-type leaf
A. Subfamily: Arabidopsis thaliana hypothetical protein Atgtp19_170

Query Match	3.74;	Score 8;	DB 2;	Length 291;
Similarity	100.0%;	Pred. No. 9.7;		
Matches	8;	Mismatches	0;	Indels
	0;	Conservative	0;	Gaps

QY 133 .PAAAPLT 140

DB 24 / PAVAROLI 234

RESOLV II
T48873
electron transfer protein

C: Species: *Fratauria* sp.
A: Variety: strain ANA-18

C:\Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jun-2000
C:\Accession: T48873

R:Murakami, S.; Takashima, A.; Takemoto, J.; Takenaka, S.; Shinke, R.; Aoki, K.
Gene 226, 189-198, 1999

A>Title: Cloning and sequence analysis of two catechol-degrading gene clusters from the
A/Reference number: Z24833

A/Accession: T46873

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-325 <MUR>

A/Cross-references: EMBL:AB009373; PIDN:BA075213.1

A/Experimental source: strain ANA-18

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 325;
Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAAPAAA 137

DB 223 AAAPAAA 230

RESULT 12

TPR domain protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C/Accession: A87300

R:Nierman, W.C.; Feidlyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Leub, M.T.; Debey, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A>Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: A87249; MUID:21173698; PMID:11259647

A/Accession: A87300

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-336 <STO>

A/Cross-references: GB:AE005673; NID:G13421571; PIDN:AAK2397.1; GSPDB:GMD0148

C/Genetics:

A/Gene: CC0410

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 336;
Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127

DB 112 AAAPAAP 119

RESULT 13

pupal cuticle protein precursor - greater wax moth

C/Species: Galleria mellonella (greater wax moth)

C/Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 19-May-2000

C/Accession: S41958

R:Kolberg, U.; Obermeier, B.; Hirsch, H.; Kelber, G.; Wolbert, P.

submitted to the EMBL Data Library, February 1994

A/Description: Expression cloning and characterization of a pupal cuticle protein cDNA C

A/Reference number: S41958

A/Accession: S41958

A/Molecule type: mRNA

A/Residues: 1-353 <KOL>

A/Cross-references: EMBL:X77514; NID:G453403; PIDN:CA054650.1; PID:G453404

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-353/Product: pupal cuticle protein #status predicted <MAT>

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 353;
Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAPAAA 126

DB 236 SAAPAAA 243

RESULT 14

T46827

phenoxycarboxylate dioxygenase [imported] - Acinetobacter lwoffii

C/Species: Acinetobacter lwoffii

C/Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 02-Sep-2000

C/Accession: T46827

R:Kim, S.I.; Leem, S.H.; Choi, J.S.; Chung, Y.H.; Kim, S.; Park, Y.M.; Ha, K.S.

submitted to the EMBL Data Library, November 1996

A/Reference number: Z24100

A/Accession: T46827

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-381 <KIM>

A/Cross-references: EMBL:U77659; PIDN:AA031770.1

A/Experimental source: strain K24

C/Superfamily: phthalate dioxygenase reductase; cytochrome-b5 reductase homology; ferred

Query Match

Best Local Similarity 100.0%; Score 8; DB 2; Length 381;
Pred. No. 12;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAAPAAA 137

DB 279 AAAPAAA 286

RESULT 15

cellular tumor antigen p53 - golden hamster

N/Alternate names: tumor-suppressor protein p53

C/Species: Mesocricetus auratus (golden hamster)

C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C/Accession: JH0633

R:Legros, Y.; McInyre, P.; Soussi, T.

Gene 112, 247-250, 1992

A>Title: The cDNA cloning and immunological characterization of hamster p53.

A/Reference number: JH0633; MUID:92210007; PMID:1555773

A/Accession: JH0633

A/Molecule type: mRNA

A/Residues: 1-356 <LEG>

A/Cross-references: GB:M75144; NID:G191414; PIDN:AA037085.1; PID:G191415

A/Experimental source: kidney, strain MP1

C/Genetics:

A/Gene: p53

C/Superfamily: cellular tumor antigen p53

C/Keywords: apoptosis; cell division control; DNA binding; homodimer; nucleus; phosf

F:179,182,241,245/Binding site: zinc (Cys, His, Cys, Cys) #status predicted

F:395/Binding site: phosphoryl-RNA (Ser) (covalent) #status predicted

Query Match

Best Local Similarity 100.0%; Score 8; DB 1; Length 396;
Pred. No. 13;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAPAAP 127

DB 66 AAAPAAP 73

Search completed: January 15, 2004, 15:30:23

Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:07:29 ; Search time 41 Seconds

(without alignments)
847.832 Million cell updates/sec

Title: US-09-923-236-2.

Perfect score: 1178
Sequence: 1 MKLLWACIVCAFARRRRF.....TAKPADEPHPSLEGANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: /SIDS1/gcgdata/geneseq/emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1178	100.0	219	22	AAB60109
2	1178	100.0	219	23	AAE27862
3	1178	100.0	219	23	ABG31608
4	1178	100.0	219	23	AAU74536
5	1178	100.0	219	24	ABU08515
6	1172	99.5	219	20	AAW30653
7	1113.5	94.5	221	22	AAI19472
8	1113.5	94.5	325	20	ABG21919
9	180	15.3	207	14	AAK40797

10	179	15.2	207	17	AAW06915
11	179	15.2	207	20	AAI32850
12	179	15.2	207	20	AAI23305
13	175.5	14.9	718	22	AAW42089
14	175.5	14.9	903	23	ABP41277
15	175.5	14.9	3177	22	AAW40303
16	174	14.8	566	22	ABE61040
17	173	14.7	538	22	AAE82806
18	173	14.7	1427	23	ABP63024
19	171.5	14.6	693	23	ABP69529
20	170.5	14.5	149	22	AAU00451
21	170	14.4	267	19	AAW81726
22	170	14.4	267	19	AAW64359
23	170	14.4	267	20	AAI39156
24	170	14.4	267	20	AAI39013
25	170	14.4	428	23	AAW50963
26	170	14.4	428	23	ABP70532
27	170	14.4	842	22	ABE66631
28	170	14.4	864	22	ABE71319
29	170	14.4	1028	22	ABU70689
30	170	14.4	3151	24	ABR39571
31	170	14.4	3173	23	ABR90743
32	170	14.4	3173	24	ABU54450
33	170	14.4	3176	24	ABR39570
34	169.5	14.4	783	19	AAW37151
35	169.5	14.4	787	19	AAW37152
36	169.5	14.4	802	19	AAW37153
37	169.5	14.4	802	22	AAU09139
38	169	14.3	299	22	ABE63276
39	168.5	14.3	550	22	ABE82807
40	168.5	14.3	706	22	ABE66633
41	165.5	14.0	763	18	AAW31852
42	164	13.9	572	18	AAW31855
43	163.5	13.9	214	17	AAE86913
44	162.5	13.8	471	22	ABE61785
45	162	13.8	1064	22	ABE66652

ALIGNMENTS

RESULT 1					
AAB60109					
ID	AAB60109 standard; Protein; 219 AA.				
XX					
AC	AAB60109;				
XX					
DT	28-MAR-2001 (first entry)				
XX					
DE	Human transport protein TPPT-29.				
XX					
KW	Human; transport protein; TPPT; transport disorder; metabolic disorder;				
KW	neurological disorder; cardiovascular disorder; reproductive disorder;				
KW	immune disorder; cancer.				
XX					
OS	Homo sapiens.				
XX					
FN	W0200078953-A2.				
XX					
PD	28-DEC-2000.				
XX					
PE	16-JUN-2000; 2000MO-US16668.				
XX					
PR	17-JUN-1999; 99US-0139923.				
PR	10-AUG-1999; 99US-0148177.				
PR	18-AUG-1999; 99US-0149357.				
PR	28-OCT-1999; 99US-0162287.				
XX					
PA	(INCY-) INCYTE GENOMICS INC.				
XX					
PI	Lai P, Yang J, Yue H, Hillman JL, Tang YT, Bandman O, Burford N;				
PI	Baughn MR, Azimzai Y, Lu DM, Au-Young J, Patterson C;				
XX					

T. cruzi Tcd anti
Tcd protein sequen
Tcd antigen of TTP
Human polypeptide
Human ovarian anti
Human polypeptide
Drosophila melanog
Human low density
Human polypeptide
Human polypeptide
Protein encoded by
M. tuberculosis Im
M. tuberculosis Im
Mycobacterium tube
M. tuberculosis an
M. tuberculosis re
Maize methyl Cpg b
Histone deacetylase
Drosophila melanog
Drosophila melanog
Human adipocyte Se
Human alpha3 colla
Human Tumour Endoc
Human tumour endot
Human alpha3 colla
Mouse neural Mena+
Mouse neural Mena+
Mouse neural Mena+
Mammalian enabled
Drosophila melanog
Rabbit low density
Drosophila melanog
Mycobacterium tube
Mycobacterium tube
Cotton fiber-speci
Drosophila melanog
Drosophila melanog

DR WPI; 2001-041424/05.
 DR N-PSDB; AAF27729.
 XX Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 PS Claim 2; Page 130; 165pp; English.
 XX The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated rPTs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 XX
 SQ Sequence 219 AA;
 Query Match 100.0%; Score 1178; DB 22; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2, 5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLLLMACIVCAVPAKRRPPPIGDDNDGHLHSLNIPVIGIRLPPPLYRPNVTP 60
 DB 1 MLLLMACIVCAVPAKRRPPPIGDDNDGHLHSLNIPVIGIRLPPPLYRPNVTP 60
 QY 61 SYPGNTYTDGLPSYPMILTSPPGFVYVHIGFPLATOLNVPPLPRGFPVPPSRFSA 120
 DB 61 STPGNTYTDGLPSYPMILTSPPGFVYVHIGFPLATOLNVPPLPRGFPVPPSRFSA 120
 QY 121 AAAPAAPPIAEBPAAAPLTATPVAAPPAAGAPVAAPVAAPVAAPVAAPVAAP 180
 DB 121 AAAPAAPPIAEBPAAAPLTATPVAAPPAAGAPVAAPVAAPVAAPVAAPVAAP 180
 QY 181 EAPVGVPAABEPSPAEPATAPKPAABEPHPSLEQANQ 219
 DB 181 EAPVGVPAABEPSPAEPATAPKPAABEPHPSLEQANQ 219
 RESULT 2
 ID AAE27862 standard; Protein; 219 AA.
 AC AAE27862;
 XX 27-DEC-2002 (first entry)
 DT 27-DEC-2002 (first entry)
 DE Human zsig63 protein.
 XX Human; secreted salivary protein; zsig63 protein; host defense protein;
 KM immune modulating factor; antipathogenic; cell-cell signalling molecule;
 KM growth factor; cytokine; growth factor hormone activity; dental carrier;
 KM infection; tooth decay; periodontal disease; gastrointestinal disease;
 KM thrush; urinary tract infection; vaginal infection; diabetes; obesity;
 KM anti-inflammatory; chronic tissue damage; lung dysfunction; testostis;
 KM gene therapy; salivary gland dysfunction; prostate gland dysfunction;
 KM forensic DNA profiling; chondrosarcoma; atherosclerosis; chromosome 4.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH 1..15
 FT Peptide /label= "Signal_peptide"
 FT 16..219
 FT Protein /note= "Mature human zsig63 protein"
 FT 14..21
 FT Region /note= "Hydrophilic region"
 FT 16..37
 FT Domain /note= "Domain 1"
 FT 17..33
 FT Region /note= "Antigenic epitope"
 FT 24..30
 FT Region /note= "Hydrophilic region"
 FT 38..126
 FT Domain

FT /note= "Domain 2"
 FT 66..73
 FT Region /note= "Antigenic epitope"
 FT 103..108
 FT Region /note= "Antigenic epitope"
 FT 124..133
 FT Region /note= "Repeat 1"
 FT 127..219
 FT Domain /note= "Domain 3"
 FT 134..138
 FT Region /note= "Repeat 2"
 FT 139..143
 FT Region /note= "Repeat 3"
 FT 144..148
 FT Region /note= "Repeat 4"
 FT 149..153
 FT Region /note= "Repeat 5"
 FT 154..158
 FT Region /note= "Repeat 6"
 FT 159..163
 FT Region /note= "Repeat 7"
 FT 164..168
 FT Region /note= "Repeat 8"
 FT 169..173
 FT Region /note= "Repeat 9"
 FT 174..178
 FT Region /note= "Repeat 10"
 FT 179..183
 FT Region /note= "Repeat 11"
 FT 184..188
 FT Region /note= "Repeat 12"
 FT 187..192
 FT Region /note= "Hydrophilic region"
 FT 189..193
 FT Region /note= "Repeat 13"
 FT 190..197
 FT Region /note= "Antigenic epitope"
 FT 194..198
 FT Region /note= "Repeat 14"
 FT 199..203
 FT Region /note= "Repeat 15"
 FT 202..215
 FT Region /note= "Antigenic epitope"
 FT 204..208
 FT Region /note= "Repeat 16"
 FT
 FT
 XX
 XX
 XX US2002090677-A1.
 XX
 XX 11-JUL-2002.
 XX
 XX 03-AUG-2001; 2001US-0923236.
 XX
 XX 17-MAR-1999; 99US-124820P.
 XX 17-MAR-2000; 2000US-0527345.
 XX
 XX (ADLER/) ADLER D A.
 XX (SHEP/) SHEPPARD P O.
 XX
 XX
 XX Adler DA, Sheppard PO;
 XX WPI; 2002-642378/69.
 XX N-PSDB; AAD45050.
 XX
 XX Novel secreted salivary polypeptide, zsig63, useful as antimicrobial
 PT agent for treating microbial infection, dental carrier, periodontal
 PT disease, thrush gastrointestinal disease, and for aiding digestion -
 PS Claim 10; Page 28; 33pp; English.
 XX
 XX The invention relates to human secreted salivary polypeptide designated
 CC as zsig63 and nucleic acid molecules encoding such polypeptides. zsig63
 CC can be used in detecting agonists and antagonists of its activity, and
 CC is also useful as a host defense polypeptide, immune modulating factor,

CC antipathogenic polypeptide, cell-cell signalling molecule, growth factor,
 CC cytokine, or as secreted extracellular matrix associated proteins with
 CC growth factor hormone activity. It is useful for treating conditions
 CC associated with pathological microbes, including bacterial, fungal and
 CC viral infections, for treating dental caries (tooth decay), periodontal
 CC disease, thrush and gastrointestinal disease, for treating urinary tract
 CC infection, vaginal infection and for preventing infection in skin and
 CC other epithelial wounds. zsig63 is useful for establishing normal
 CC microflora and protect against pathogenic colonisation and invasion, for
 CC treating chronic tissue damage e.g. damage in extremities associated
 CC with diabetes and useful as anti-inflammatory agents. It is useful as a
 CC marker of lung dysfunction, salivary gland dysfunction, or dysfunction of
 CC prostate gland. It is also therapeutically useful for aiding digestion.
 CC Polynucleotides of the invention are used in gene therapy for increasing
 CC or inhibiting zsig63 activity, for detecting abnormalities on human
 CC chromosome 4 associated with disease or other human traits and as
 CC diagnostics in forensic DNA profiling. Sequences of the invention are
 CC useful for stimulating proliferation or differentiation of cardiac
 CC myocytes, for proliferation or differentiation of adipocytes and for
 CC inhibiting chondrosarcomas, atherosclerosis, restenosis and obesity.
 CC The present sequence is human zsig63 protein. zsig63 gene is located
 CC on chromosome 4.

CC Sequence 219 AA:

Query Match 100.0%; Score 1176; DB 23; Length 219;
 Best Local Similarity 100.0%; Pred. No. 2.5e-85;
 Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLLIMACTIVCAAPKRRPPIGDDNDGHPILPSINITYGRNLPPPLYRRPVNV 60
 Db 1 MKLLIMACTIVCAAPKRRPPIGDDNDGHPILPSINITYGRNLPPPLYRRPVNV 60
 QY 61 SYPGNTYDTGLPSYPIWLTSPGFPYVYHNGFPLATOLNPPLPPRPGFPVPSRPFSA 120
 Db 61 SYPGNTYDTGLPSYPIWLTSPGFPYVYHNGFPLATOLNPPLPPRPGFPVPSRPFSA 120
 QY 121 AAAPAPPIAABPAAAPLTLTPVAABPAGAPVAABPAAAPVGAEPAAAPVAABPAA 180
 Db 121 AAAPAPPIAABPAAAPLTLTPVAABPAGAPVAABPAAAPVGAEPAAAPVAABPAA 180
 QY 181 EAPVGEPAABEPSPAPAPATKPAAPRPHPSLEQANQ 219
 Db 181 EAPVGEPAABEPSPAPAPATKPAAPRPHPSLEQANQ 219

RESULT 3

ABG31608 standard; Proteins; 219 AA.

ABG31608;

15-NOV-2002 (first entry)

Human secreted salivary protein zsig63.

Human; secreted salivary protein; zsig63; immunogen; zsig63-cytokine;
 antibody-cytokine; in vivo killing; pathological microbes; bacteria;
 fungal; viral; infection; salivary gland; anti-microbial; dental caries;
 tooth decay; periodontal disease; thrush; gastrointestinal disease;
 urinary tract infection; vaginal infection; skin infection; microflora;
 epithelial wound; pathogenic colonisation; invasion; pro-inflammatory;
 chronic tissue damage; vascular system; diabetes; anti-inflammatory;
 incompetent immune system; AIDS; acquired immunodeficiency syndrome;
 chemotherapy; radiation treatment; lung infection; cystic fibrosis;
 digestion; chromosome 4.

Homo sapiens.

Key Location/Qualifiers

Region 1..15 /note= "Fusion protein peptide, specifically claimed in claim 18"

FT Region 1..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 9..204 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 14..19 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..21 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..37 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 16..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 17..33 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..29 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 24..33 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 25..30 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 38..126 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 66..73 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 103..108 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 109..215 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 124..133 /label= Repeat_1
 FT 127..219 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 134..138 /label= Repeat_2
 FT 139..143 /label= Repeat_3
 FT 144..148 /label= Repeat_4
 FT 149..153 /label= Repeat_5
 FT 154..158 /label= Repeat_6
 FT 159..163 /label= Repeat_7
 FT 164..168 /label= Repeat_8
 FT 169..173 /label= Repeat_9
 FT 174..178 /label= Repeat_10
 FT 179..183 /label= Repeat_11
 FT 184..188 /label= Repeat_12
 FT 187..192 /note= "Antigenic peptide, specifically claimed in
 FT claim 15"
 FT Region 189..193

FT Region /label= Repeat_13
FT 190..197
FT /note= "Antigenic peptide, specifically claimed in
FT claim 15"
FT Region 194..198
FT /label= Repeat_14
FT Region 199..203
FT /label= Repeat_15
FT Region 202..215
FT /note= "Antigenic peptide, specifically claimed in
FT claim 15"
FT Region 204..208
FT /label= Repeat_16
PN US2002081701-A1.
PD 27-JUN-2002.
PF 03-AUG-2001; 2001US-0922480.
PR 17-MAR-1999; 99US-124820P.
PR 17-MAR-2000; 2000US-0527345.
XX (ADLER/) ADLER D A.
XX (SHEP/) SHEPPARD P O.
XX Adler DA, Sheppard PO;
XX WPI: 2002-635468/68.
XX N-PSDB: ABS52633, ABS52634.
XX Claim 10; Page 28; 33pp; English.
XX The present invention relates to a new secreted salivary protein, zsig63.
XX The invention is useful for detecting in a test sample, the presence of
XX an antagonist or agonist of zsig63 protein activity. The invention is
XX also useful as an immunogen for producing an antibody to zsig63
XX polypeptide. zsig63-cytokine fusion proteins or antibody-cytokine fusion
XX protein are useful for enhancing in vivo killing of target tissues.
XX Pharmaceutical composition comprising purified zsig63 polypeptide are
XX useful in the treatment of conditions associated with pathological
XX microbes, including bacterial, fungal and viral infections. High
XX expression of zsig63 in salivary gland suggests that anti-microbial
XX polypeptides are useful for treatment of dental caries (tooth decay),
XX periodontal disease, thrush and gastrointestinal disease. Other
XX applications can be used in urinary tract infections, vaginal infections,
XX prevention of infection in skin and other epithelial wounds. The
XX polypeptides can be used to establish normal microflora and protect
XX against pathogenic colonisation and invasion. The invention is useful
XX when pro-inflammatory activity is desired. Applications for
XX such pro-inflammatory activity include the treatment of chronic tissue
XX damage, particularly in areas having a limited or damaged vascular system
XX e.g., damage in extremities associated with diabetes. Antagonists to
XX zsig63 polypeptides may be useful as anti-inflammatory agents. The
XX invention is useful for the treatment of patients having incompetent
XX immune system, such as AIDS (acquired immunodeficiency syndrome) patients
XX or individuals that have undergone chemotherapy, radiation treatment. The
XX invention is also useful for the treatment of lung infections associated
XX with cystic fibrosis and its agonists or antagonists are useful for
XX aiding digestion. The present amino acid sequence represents the human
XX secreted salivary protein zsig63 of the invention. This sequence is
XX encoded by the human zsig63 gene located on chromosome 4.
XX Sequence 219 AA;

QY	1	MLKLLMACVCAVCAFRKKRRFPPIGGDDNDGDCHPLHPSLNTPIYGRINLPPLYYRPVNTVP	60
Db	1	MKLLIMACIVCAFRKKRRFPPIGGDDNDGDCHPLHPSLNTPIYGRINLPPLYYRPVNTVP	60
QY	61	SYPGNTYTDYGLPSYPMILITLSBGPYYVNHINGPFLATQLVNLPPLPRGPFVPVPSRFSFA	120
Db	61	SYPGNTYTDYGLPSYPMILITLSBGPYYVNHINGPFLATQLVNLPPLPRGPFVPVPSRFSFA	120
QY	121	AAAPAAPPIAAEBPAAAALPLTATPVAAEBPAAAGAPVAAEBPAAAPVGAAPAAAPVAAEPAA	180
Db	121	AAAPAAPPIAAEBPAAAALPLTATPVAAEBPAAAGAPVAAEBPAAAPVGAAPAAAPVAAEPAA	180
QY	181	BAVGVGPAAEBSPBABPATAKPAAPEHPSPSLEQANQ	219
Db	181	BAVGVGPAAEBSPBABPATAKPAAPEHPSPSLEQANQ	219

RESULT 4	
AAU74536	
ID	AAU74536 standard; Protein; 219 AA.
XX	
AC	AAU74536;

DT	23-APR-2002 (first entry)
XX	
DE	Human zsig63 polypeptide.
XX	
KW	Human; zsig63; chromosome 4q12-4q13; salivary protein; antimicrobial;
KW	microbial infection; tooth decay; periodontal disease; thrush; emphysema;
KW	gastrointestinal disease; urinary tract infection; vaginal infection;
KW	skin infection; epithelial wound; chronic tissue damage; cystic fibrosis;
KW	acquired immunodeficiency syndrome; AIDS; lung infection; sarcoidosis;
KW	chronic bronchitis; gene therapy; protein therapy.
XX	
OS	Homo sapiens.
XX	
PN	US6331413-B1.
XX	
PD	18-DEC-2001.
XX	
PP	17-MAR-2000; 2000US-0527345.
XX	
PR	17-MAR-1999; 99US-124820P.
XX	
PA	(ZYMO) ZYMOGENETICS INC.
XX	
PI	Adler DA, Shepard PO;
DR	WPI; 2002-096707/13.
XX	
DR	N-PSDB; AAS20591.
XX	
PT	Polynucleotides encoding salivary proteins useful as anti-microbial
PT	agents -
PS	Claim 1; Column 49-52; 29pp; English.
XX	
CC	The invention relates to a polynucleotide derived from the 4q12-4q13
CC	region of human chromosome 4 and encoding a zsig63 polypeptide, a
CC	secreted salivary protein with anti-microbial activity. Due to their
CC	microbial activity, the sequences can be used in the study of microbial
CC	infections, e.g. for recombinant production of anti-microbial proteins.
CC	The sequences can be used in the treatment of tooth decay, periodontal
CC	disease, thrush, gastrointestinal disease, urinary tract infections,
CC	vaginal infections, skin infections, epithelial wounds, chronic tissue
CC	damage, acquired immunodeficiency syndrome (AIDS), cystic fibrosis, lung
CC	infections, sarcoidosis, emphysema and chronic bronchitis. This sequence
CC	represents human zsig63.

Query Match	100.0%	Score 1178;	DB 23;	Length 219;
Best Local Similarity	100.0%	Pred. No. 2.5e-85;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Query Match	100.0%	Score 1178;	DB 23;	Length 219;
Best Local Similarity	100.0%	Pred. No. 2.5e-85;		
Matches 219; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

[illegible]

PT		/note= "Specifically claimed in claim 15"
PT	Peptide	187..192 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
FT	Peptide	24...33 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Peptide	17...33 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Peptide	66...73 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Peptide	103...108 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Peptide	190..197 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Peptide	202...215 /label= Immunogenic peptide
FT		/note= "Specifically claimed in claim 15"
PT	Region	124...133 /label= Repeat_1
FT		134..138 /label= Repeat_2
PT	Region	139..143 /label= Repeat_3
FT		144...148 /label= Repeat_4
PT	Region	149..153 /label= Repeat_5
FT		154..158 /label= Repeat_6
PT	Region	159..163 /label= Repeat_7
FT		164..168 /label= Repeat_8
PT	Region	169..173 /label= Repeat_9
FT		174..178 /label= Repeat_10
PT	Region	179..183 /label= Repeat_11
FT		184..188 /label= Repeat_12
PT	Region	189..193 /label= Repeat_13
FT		194..198 /label= Repeat_14
PT	Region	199..203 /label= Repeat_15
FT		204..208 /label= Repeat_16
XX	US2002173027-A1.	
PX		
PD	21-NOV-2002.	
XX		
PR	03-AUG-2001, 2001US-0922469.	
XX		
PR	17-MAR-1999; 99US-124820P.	
PR	17-MAR-2000; 2000US-0527345.	
PA	(ADLER/) ADLER D A.	
XX	(SHEP/) SHEPPARD P O.	
XX		
PI	Adler DA, Sheppard PO;	
OR	WPI, 2003-328428/31.	

DR N-PSDB; ABX93594.
XX
XX Novel isolated zsig63 polypeptide, member of the adhesin family, useful
PT for treating dental caries, periodontal disease, thrush,
PT gastrointestinal disease, urinary tract infections, vaginal infections,
PT skin infections
XX
XX Claim 10; Page 27-28; 32pp; English.
XX
XX The invention relates to an isolated zsig63 polypeptide comprising at
CC least 90% identity to an amino acid sequence which comprises domain 1 of
CC zsig63, domain 2, domain 3, mature zsig63 and full length zsig63. Also
CC included are the polynucleotide encoding zsig63, a zsig63 expression
CC vector, a cultured cell comprising the vector and expressing the protein,
CC a DNA encoding a fusion protein (comprising amino acids 1-15, 16-37,
CC 38-126, 127-219 or 16-219 of zsig63 and an additional protein), using a
CC zsig63 reporter gene construct to identify zsig63 agonists, and
CC producing an anti-zsig63 antibody using zsig63 immunogenic peptides,
CC zsig63 is useful for detecting in a test sample, the presence of
CC antagonist of zsig63 protein activity. Zsig63 has antimicrobial
CC activity and since exhibits high expression in salivary gland, can be
CC used for treating dental caries, periodontal disease, thrush, and
CC gastrointestinal disease, urinary tract infections, vaginal infections,
CC skin infections and other epithelial wounds. The polypeptides can be
CC used to establish normal microflora and protect against pathogenic
CC colonization and invasion. Zsig63 can also be used for providing
CC pro-inflammatory activity for treating chronic, tissue damage
CC particularly in areas having limited or damaged vascular system, e.g.
CC in diabetes, and for treating immunocompromised AIDS patients or in
CC individuals that have undergone chemotherapy, radiation treatment, for
CC treating lung infections e.g. in cystic fibrosis. Detection of zsig63
CC polypeptide at relatively high levels in the trachea may indicate that
CC such polypeptides may serve as a marker of lung dysfunction. Zsig63 is
CC also useful in diagnosing conditions associated with salivary gland or
CC lung dysfunction including salivary gland carcinoma, pneumocystis carinii
CC infection, emphysema, chronic bronchitis, prostate dysfunctions such
CC as prostate adenocarcinoma, aiding digestion, and as components of
CC defined cell culture media and may be used to replace serum that is
CC commonly used in culture. The DNA is useful in gene therapy applications
CC to increase or inhibit zsig63 activity, and for detecting abnormalities
CC on human chromosome 4 (e.g. 4q12-4q13, associated with dentinogenesis
CC imperfecta, and dentin dysplasia type II). Zsig63 is an adhesin family
CC member. The present sequence represents human zsig63.
XX
XX
SQ Sequence 219 AA;
Query Match 100.0%; Score 1178; DB 24; Length 219;
Best Local Similarity 100.0%; Pred. No. 2, 5e-85;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLMACTVCVAFARKRRFPPIGDDNDGHLPSLNIPIGIRLPPPLYRPNVTP 60
DB 1 MKLLMACTVCVAFARKRRFPPIGDDNDGHLPSLNIPIGIRLPPPLYRPNVTP 60
QY 61 SYPGNTYTDGLPSYPMILTSFGPFVYHIRGFPPLATQLVNPPPLPRGPFVPPSRFSA 120
DB 61 SYPGNTYTDGLPSYPMILTSFGPFVYHIRGFPPLATQLVNPPPLPRGPFVPPSRFSA 120
QY 121 AAPAAPPIAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAEPAAPVAAEPAA 180
DB 121 AAPAAPPIAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAEPAAPVAAEPAA 180
QY 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219
DB 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219

RESULT 6
AAW30653
ID AAW30653 standard; Protein; 219 AA.
XX
AC AAW30653;
XX

DT 12-APR-1999 (first entry)
XX
XX Human secreted protein clone cp16 1 protein.
DE
XX Human; secreted protein; nutritional activity; cytokine; vaccine;
KW cell proliferation; differentiation; immune stimulation; suppression;
KW haematopoiesis regulation; tissue growth; activin; inhibin; chemotactic;
KW chemokine; haemostatic; thrombolytic; anti-inflammatory; gene therapy;
KW tumour invasion suppression; tumour inhibition.
XX
XX Homo sapiens.
OS
XX WO9901466-A1.
XX
XX 14-JAN-1999.
XX
XX 01-JUL-1998; 98WO-US13813.
XX
XX 27-OCT-1997; 97US-0958304.
XX
XX 02-JUL-1997; 97US-0887195.
XX
XX (GENV) GENETICS INST INC.
XX
XX Agostino M; Jacobs K, Lavallee BR, McCoy JM, Racie LA;
PI Spaulding V, Treacy M;
PI
XX MPI; 1999-105994/09.
XX
XX N-PSDB; AAW80740.
XX
XX New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult testes, adult brain, foetal kidney,
PT adult salivary gland, or adult blood cDNA libraries, useful as, e.g.
PT potential vaccines
XX
XX Claim 24; Page 71-72; 107pp; English.
XX
XX The present sequence represents a human secreted protein from clone
CC cp16 1, deposited as ATCC 98482. Human secreted protein clone
CC polynucleotides and proteins are predicted to have biological
CC activities which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals. Suggested
CC activities include nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating (e.g. as
CC vaccine) or suppressing activity, haematopoiesis regulating activity,
CC tissue growth activity, activin/inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, anti-inflammatory activity, cadherin/tumour invasion
CC suppressor activity, and tumour inhibition activity. The polynucleotides
CC are also stated to be useful for gene therapy.
XX
XX
SQ Sequence 219 AA;
Query Match 99.5%; Score 1172; DB 20; Length 219;
Best Local Similarity 99.1%; Pred. No. 7, 4e-85;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKLLMACTVCVAFARKRRFPPIGDDNDGHLPSLNIPIGIRLPPPLYRPNVTP 60
DB 1 MKLLMACTVCVAFARKRRFPPIGDDNDGHLPSLNIPIGIRLPPPLYRPNVTP 60
QY 61 SYPGNTYTDGLPSYPMILTSFGPFVYHIRGFPPLATQLVNPPPLPRGPFVPPSRFSA 120
DB 61 SYPGNTYTDGLPSYPMILTSFGPFVYHIRGFPPLATQLVNPPPLPRGPFVPPSRFSA 120
QY 121 AAPAAPPIAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAEPAAPVAAEPAA 180
DB 121 AAPAAPPIAEPAAAAPLTATPVAAEPAAAGAPVAAEPAAEAVGAEPAAPVAAEPAA 180
QY 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219
DB 181 EAPVGEPAABEPSPAPATAPAAPEPHSPSLDAQ 219

RESULT 7
AA19472
ID AA19472 standard; Protein; 221 AA.
XX
AC AA19472;
XX
DT 14-JUL-1999 (first entry)
XX
DE Amino acid sequence of a human secreted protein.
XX
KW Human secreted protein; cancer; tumour; neurodegenerative disorder;
KW developmental abnormality; fetal deficiency; blood disorder; leukemia;
KW immune system disease; autoimmune disease; hepatic disease;
KW renal disease; lymphoma; inflammation; allergy; ischemic shock;
KW Alzheimer's; cognitive disorder; schizophrenia; prostate disease;
KW obesity; osteoclast; osteoporosis; arthritis; malignancy; testes disease;
KW lung disease; thymus disease; digestive disorder; endocrine disorder;
KW infection; AIDS.
XX
OS Homo sapiens.
XX
PN MO9922243-A1.
XX
PD 06-MAY-1999.
XX
PE 23-OCT-1998; 98WO-US22376.
XX
PR 24-OCT-1997; 97US-0063387.
PR 24-OCT-1997; 97US-0062784.
PR 24-OCT-1997; 97US-0063088.
PR 24-OCT-1997; 97US-0063089.
PR 24-OCT-1997; 97US-0063090.
PR 24-OCT-1997; 97US-0063091.
PR 24-OCT-1997; 97US-0063092.
PR 24-OCT-1997; 97US-0063097.
PR 24-OCT-1997; 97US-0063098.
PR 24-OCT-1997; 97US-0063099.
PR 24-OCT-1997; 97US-0063100.
PR 24-OCT-1997; 97US-0063101.
PR 24-OCT-1997; 97US-0063109.
PR 24-OCT-1997; 97US-0063110.
PR 24-OCT-1997; 97US-0063111.
PR 24-OCT-1997; 97US-0063148.
PR 24-OCT-1997; 97US-0063386.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Brewer LA, Carter KC, Duan DR, Ebner R, Endress GA;
PI Feng P, Florence C, Florence KA, Greene JM, Janat F;
PI Kaye H, Lafleur DW, Moore PA, Ni J, Olsen HS, Rosen CA;
PI Ruben SM, Shi Y, Soppet DR, Wei Y, Young P;
XX
DR WPI: 1999-303069/25.
DR N-PSDB; AAX61352.
XX
PT New isolated human genes and the secreted polypeptides they encode
XX
PS Claim 11; Page 401-402; 546pp; English.
XX
CC The specification describes cDNA sequences (AAX61322-X61470) encoding
CC human secreted proteins (AA19442-Y19590). The polynucleotides and their
CC corresponding secreted polypeptides are useful for preventing, treating
CC or ameliorating medical conditions, e.g. by protein or gene therapy.
CC Pathological conditions can also be diagnosed by determining the amount
CC of the polypeptides in a sample or by determining the presence of a
CC mutations in the polynucleotides. Specific uses are described for each
CC of the polynucleotides, based on which tissues they are most highly
CC expressed in, and include developing products for the diagnosis or
CC treatment of cancer, tumours, neurodegenerative disorders, developmental
CC abnormalities and fetal deficiencies, blood disorders, leukemias,
CC diseases of the immune system, autoimmune diseases, hepatic and renal
CC diseases, lymphomas, inflammation, allergies, ischemic shock, Alzheimer's
CC and cognitive disorders, schizophrenia, prostate diseases, obesity,

CC disorders involving osteoclasts such as osteoporosis, arthritis or
CC malignancies, diseases of testes, lung or thymus, digestive/endocrine
CC disorders, infections and AIDS. The polypeptides are also useful for
CC identifying their binding partners.
XX
SQ Sequence 221 AA;
XX
Query Match 94.5%; Score 1113.5; DB 20; Length 221;
Best Local Similarity 95.0%; Pred. No. 3.1e-80;
Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;
XX
QY 1 MKLLMACIVCAVAPAKRRPPTGDDNDGHLHPSLNIPIYGRNLPPPLYRPNVTVP 60
DB 1 MKLLMACIVCAVAPAKRRPPTGDDNDGHLHPSLNIPIYGRNLPPPLYRPNVTVP 60
QY 61 SYQNTTDTGDLSPYPIILSPGPPYYHTRGPPLAQLVNPPPPRGPVPVPSRPFSA 120
DB 61 SYQNTTDTGDLSPYPIILSPGPPYYHTRGPPLAQLVNPPPPRGPVPVPSRPFSA 120
QY 121 AAAPAPPIAEPAPAAAPLTATPVAAEPAPAGAPVAPAPAP- VQAPAPAPVAAEP 179
DB 121 AAAPAPPIAEPAPAAAPLTATPVAAEPAPAGAPVAPAPAP- VQAPAPAPVAAEP 180
QY 180 AEAAPGVAPAEBSPPAPATKAPADBPSPSLEQANQ 219
DB 181 AEAAPGVAPAEBSPPAPATKAPADBPSPSLEQANQ 220
XX
RESULT 8
ABG21919
ID ABG21919 standard; Protein; 325 AA.
XX
AC ABG21919;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21910.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS86106.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID NO 52278; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 325 AA;

Query Match 15.6%; Score 183.5; DB 22; Length 325;
 Best Local Similarity 34.3%; Pred. No. 9.5e-07;
 Matches 57; Conservative 8; Mismatches 78; Indels 23; Gaps 6;

QY 48 PPPLYRPPVNTVSIGNTYTDGLSPYEWILSPGEPVYTHRGPIATQNVPLPR 107
 DB 45 PCPSRYCPQMTPPPP-----PAPPLPPPPPP-----APAPPLPPAPSP- 87
 QY 108 GPFVPPSPSPFAAAPPAPPIAAPPAAAPLTATPVAAEPAGAPVAAEPAAEPVGA 167
 DB 88 --PHAPPPQALPPPPALPPPP--ASPPLPPALSPPLPAPPSAPSPAPSPPPAP 143
 QY 168 PAAPAPVAAEPAAEPVGAEPAAEPSPAPATAPAPAPAPSP 212
 DB 144 PPSPPAPPSAPSPAPPLPPAPSPAPSPAPPPAPPPAPPPAPPP-PPSP 188

RESULT 9

AA040797
 ID AAR40797 standard; Protein; 207 AA.

AC AAR40797;
 DT 25-MAR-2003 (updated)
 DT 08-FEB-1994 (first entry)
 DE Sequence of a 10-amino acid repetitive sequence of an antigenic
 DE peptide domain encoded by an insert of clone Tcd.
 XX
 XX Epitope repeat; antigen; Chagas disease; TCD.
 XX
 OS Trypanosoma cruzi.

XX
 FH Key Location/Qualifiers
 FT Misc-difference 42
 FT /label= Degeneracy in repeat unit
 FT /note= "see also Aas 88,108,158,162"

XX WO9316199-A1.

XX 19-AUG-1993.

XX 11-FEB-1993; 93WO-US01231.

XX 14-FEB-1992; 92US-0836642.

XX (REBD/) REED S G.

XX Reed SG;

XX WPI; 1993-272900/34.

XX N-PSDB; AA040797.

XX Antigenic peptide domain of trypanosoma cruzi - used to diagnose
 PT or immunise against infection, and screen blood supplies

PS Disclosure; Figure 1; 16pp; English.

XX Clone Tcd encodes a 10-amino acid repetitive sequence present
 CC in 20.5 copies with minor degeneracies present in 5 posns.
 CC The predicted mol. wt. of recombinant unglycosylated Tcd antigen
 CC is 36.3KD. Analysis of DNA from seven geographically diverse
 CC T. cruzi isolates indicates that Tcd gene sequence was conserved.
 CC A 636 bp fragment of clone Tcd was used to probe DNA from several
 CC other protozoan parasites of humans. However, it did not hybridize
 CC to T. brucei, Leishmania chagasi, L. amazonensis, L. donovani
 CC or T. rangeli.
 CC (Updated on 25-MAR-2003 to correct PN field.)

SO Sequence 207 AA;

Query Match 15.3%; Score 180; DB 14; Length 207;
 Best Local Similarity 38.4%; Pred. No. 1.1e-06;
 Matches 43; Conservative 11; Mismatches 46; Indels 12; Gaps 2;

QY 114 PPRFSAAPAPAP-----PIAAEPAAAPLTATPVAAEPAGAPVAAEPAAEPV 164
 DB 85 PAESKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPK 144
 QY 165 GAEPAAEPVAAEPAAEPVGAEPAAEPSPAPATAPAPAPSP 213
 DB 145 PAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPK 196

RESULT 10

AA066915
 ID AA066915 standard; Protein; 207 AA.

AC AA066915;
 DT 02-APR-1997 (first entry)
 DT T. cruzi Tcd antigen.
 DE Tcd; diagnosis; infection; antigen; Chagas disease.
 XX
 XX Trypanosoma cruzi.

XX
 FH Key Location/Qualifiers
 FT Region 6..20
 FT /label= Antigenic epitope
 FT /note= "Claim 3, page 38"

XX WO9629605-A2.

XX 26-SEP-1996.

XX 12-MAR-1996; 96WO-US03380.

XX 14-MAR-1995; 95US-0403379.

XX (CORI-) CORIXA CORP.

XX Reed SG;

XX WPI; 1996-485445/48.

XX N-PSDB; AAT46149.

XX Detecting and preventing T. cruzi infection - using polypeptide(s)
 PT or antibodies contg. or reactive with antigen epitope(s) of T. cruzi
 PT proteins

XX Disclosure; Fig 4; 59pp; English.

XX The Tcd antigen (AA066915) of Trypanosoma cruzi includes an
 CC antigenic epitope (see also AA066917) that can be utilized, pref.
 CC with other T. cruzi epitopes (see also AA066914-19), in immunoassays
 CC for T. cruzi infection (Chagas' disease) and in vaccine compns.

SQ Sequence 207 AA;

Query Match 15.2%; Score 179; DB 17; Length 207;
Best Local Similarity 41.0%; Pred. No. 1.3e-06;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAP 173

DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63

QY 174 VAAEPAAEAPVGEPAEESP-----AEPTAKPAEPHPSPS 213

DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 11

ID AAY32850 standard; Protein; 207 AA.

XX AAY32850;

DT 29-OCT-1999 (first entry)

XX TCD protein sequence.

XX TCD repeat sequence; 35kD T. cruzi homologue; immune system disorder;

XX eukaryotic ribosomal protein L19E; infection; detection; diagnosis;

XX Chagas' disease; Leishmania infection.

XX Trypanosoma cruzi.

XX US942403-A.

XX 24-AUG-1999.

XX 15-SEP-1997; 97US-0929414.

XX 14-MAR-1995; 95US-0403379.

XX (CORI-) CORIXA CORP.

XX Houghton R, Reed SG, Skeiky YAW;

XX WPI; 1999-517419/43.

XX Trypanosoma cruzi antigenic homologue of eukaryotic ribosomal

XX protein L19E, useful for screening assays to detect T. cruzi

XX infection

XX Example 3; Fig 4; 26pp; English.

XX This sequence is the Trypanosoma cruzi Tcd protein. Antigenic fragments

XX of TCD and fragments of the 35 kD Trypanosoma cruzi homologue (Tcd) of

XX the eukaryotic ribosomal protein L19E can be used in the method of the

XX invention for detecting T. cruzi infection in a biological sample, which

XX comprises: (a) contacting a biological sample with a first polypeptide

XX comprising the sequence shown in AAY32850; (b) contacting the biological

XX sample with a second polypeptide comprising the TCD sequence shown in

XX AAY32850 or AAY32841; and (c) detecting the presence of antibodies that

XX bind to at least one of the polypeptides, indicating T. cruzi infection.

XX The methods and polypeptides may be used to detect T. cruzi (which

XX causes Chagas' disease and a variety of immune system disorders) and/or

XX Leishmania infection in individuals and blood supplies. The compounds

XX and methods may also be used to protect against T. cruzi infection. The

XX N-terminal region of TCD (residues 1-136) cross-reacts with

XX anti-Leishmania antibodies and can be specifically excluded from the

XX antigenic polypeptide to avoid such cross-reactivity.

SQ Sequence 207 AA;

Query Match 15.2%; Score 179; DB 20; Length 207;

Best Local Similarity 41.0%; Pred. No. 1.3e-06;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSAAAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAPPAAP 173

DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63

QY 174 VAAEPAAEAPVGEPAEESP-----AEPTAKPAEPHPSPS 213

DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 12

ID AAY23305 standard; Peptide; 207 AA.

XX AAY23305;

DT 31-AUG-1999 (first entry)

XX TCD antigen of Trypanosoma cruzi.

XX Trypanosoma cruzi epitope; Trypanosoma cruzi infection; antigen;

XX vaccine; Chagas' disease; TCD antigen.

XX Trypanosoma cruzi.

XX WO9931246-A1.

XX 24-JUN-1999.

XX 04-DEC-1998; 98WO-US25871.

XX 18-DEC-1997; 97US-0993674.

XX (CORI-) CORIXA CORP.

XX Houghton RL, Lodes MJ, McNeill PD, Reed SG, Skeiky YAW;

XX Smith JM;

XX WPI; 1999-405035/34.

XX New isolated Trypanosoma cruzi epitopes

XX Disclosure; Page 83-84; 103pp; English.

XX The specification describes new Trypanosoma cruzi epitopes. A method

XX for detecting Trypanosoma cruzi infection in a biological sample

XX comprises contacting the sample with a polypeptide comprising

XX an epitope of a TC antigen, or a variant of the antigen that differs

XX only in conservative substitutions and/or modifications and detecting

XX the presence of antibodies that bind to the polypeptide in the sample,

XX thereby detecting TC infection. The TC polypeptides can be used in

XX vaccines for inducing protective immunity against Chagas' disease in

XX a patient. The polypeptides and antibodies can also be used for detecting

XX TC infection. The present sequence represent a TCD antigen, which

XX can be used in the above assay to improve sensitivity.

SQ Sequence 207 AA;

Query Match 15.2%; Score 179; DB 20; Length 207;

Best Local Similarity 41.0%; Pred. No. 1.3e-06;

Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurologic disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pct_sequences.

XX
 XX Sequence 903 AA;
 SQ

Query Match 14.9%; Score 175.5; DB 23; Length 903;
 Best Local Similarity 32.3%; Pred. No. 1,2e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIYGRNLP---PPLYRPVNTVPSYPGNTYTDGLPSYPMILTSPPGPPYYHNR 91
 DB 579 HKQVNVNNVTSSPTSNPVTTKKVT---KPVTTTKP-----VTTTKP----- 622
 QY 92 GFLPLATQLNVPLPRGPFVPPSPRFSAAAPAA-APPIAAEPAAAAPLTATVAAEPAA 150
 DB 623 ---VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKTATVRP----- 658
 QY 151 GAPVAAEPAAEA-PVGAEPAAEAAPVAAEPAAEPVGEPAABEPSPAPATKAPAAEPH 209
 DB 659 --PVAVKPATAAKPVAAKPAAVRPPAA--AAAKPVATKEVPBPQAAKPAATKPAATTKPM 714
 QY 210 PPSPLS 215
 DB 715 VKMSRE 720

RESULT 15
 AAM40303
 ID AAM40303 standard; Protein; 3177 AA.
 XX
 AC AAM40303;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 3448.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0489725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.

XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
 PI Wang J, Wang Z, Wehman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI; 2001-442253/47.
 DR N-PSDB; AA159459.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 6; SEQ ID NO 3448; 10078bp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC assays for receptor activity, cancer diagnosis and therapy, drug screening,
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX
 SQ Sequence 3177 AA;
 XX

Query Match 14.9%; Score 175.5; DB 22; Length 3177;
 Best Local Similarity 32.3%; Pred. No. 4.4e-05;
 Matches 60; Conservative 16; Mismatches 61; Indels 49; Gaps 9;

QY 35 HPSLNIPIYGRNLP---PPLYRPVNTVPSYPGNTYTDGLPSYPMILTSPPGPPYYHNR 91
 DB 2853 HKQVNVNNVTSSPTSNPVTTKKVT---KPVTTTKP-----VTTTKP----- 2896
 QY 92 GFLPLATQLNVPLPRGPFVPPSPRFSAAAPAA-APPIAAEPAAAAPLTATVAAEPAA 150
 DB 2897 ---VTIINQPSVKP-----AAAKPAPAKPVAAKPVATKTATVRP----- 2932
 QY 151 GAPVAAEPAAEA-PVGAEPAAEAAPVAAEPAAEPVGEPAABEPSPAPATKAPAAEPH 209
 DB 2933 --PVAVKPATAAKPVAAKPAAVRPPAA--AAAKPVATKEVPBPQAAKPAATKPAATTKPM 2988
 QY 210 PPSPLS 215
 DB 2989 VKMSRE 2994

Search completed: January 15, 2004, 15:13:23
 Job time : 43 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:10:34 ; Search time 35 Seconds

(without alignments)
1614.672 Million cell updates/sec

Title: US-09-923-236-2

Sequence: 1 MKLLMAGTICVAFARRRF.....TAKPAEPHPSPLEQANQ 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

SPTRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	359.5	30.5	344	11 Q9D5J3	Q9D5J3 mus musculus
2	224.5	19.1	3179	12 Q8V2A4	Q8V2A4 human Herpe
3	221	18.8	1100	2 Q9X6J9	Q9X6J9 klebsiella
4	214.5	18.2	1480	10 Q9LIE8	Q9LIE8 arabidopsis
5	214	18.2	3105	12 Q8U2I9	Q8U2I9 cercopithec
6	206.5	17.5	607	5 Q17112	Q17112 babesia bov
7	204.5	17.4	1070	2 Q8GFS8	Q8GFS8 citrobacter
8	204	17.3	796	5 Q96579	Q96579 trypanosoma
9	199	16.9	391	5 Q26892	Q26892 trypanosoma
10	194.5	16.5	722	5 Q26893	Q26893 trypanosoma
11	193	16.4	3534	12 Q39266	Q39266 equine herp
12	188.5	16.0	1108	2 Q49542	Q49542 mycoplasma
13	188	15.8	839	16 Q9RX57	Q9RX57 deinococcus
14	186.5	15.8	395	16 Q9RT44	Q9RT44 mus musculus
15	186.5	15.8	616	11 Q99X31	Q99X31 mus musculus
16	183	15.5	744	10 Q65375	Q65375 arabidopsis

17	182	15.4	386	10 Q9FPQ5	Q9FPQ5 chlamydomon
18	181	15.4	928	12 Q9IMX9	Q9IMX9 cercopithec
19	181	15.4	2657	11 Q88493	Q88493 mus musculus
20	178.5	15.2	243	10 Q9XIV1	Q9XIV1 cucumis sat
21	177.5	15.1	598	16 Q8VKN7	Q8VKN7 mycobacteri
22	177	15.0	1315	10 Q9SEPW0	Q9SEPW0 zea mays (m
23	176.5	15.0	3084	12 Q8U2I1	Q8U2I1 pseudoraba
24	176	14.9	959	16 Q8XRH0	Q8XRH0 ralestonia s
25	175	14.9	857	16 Q98IX2	Q98IX2 rhizobium l
26	174.5	14.8	979	4 Q8N421	Q8N421 homo sapien
27	174.5	14.8	1611	3 Q42854	Q42854 schizosach
28	174	14.8	236	10 Q9LM00	Q9LM00 pinus taeda
29	174	14.8	503	12 Q39779	Q39779 equine herp
30	173.5	14.7	786	10 Q48809	Q48809 arabidopsis
31	173	14.7	1174	4 Q94854	Q94854 homo sapien
32	171.5	14.6	438	16 Q9A2W5	Q9A2W5 cauliobacter
33	171.5	14.6	616	4 Q9H6K5	Q9H6K5 homo sapien
34	171	14.5	544	12 Q89392	Q89392 parametium
35	171	14.5	1340	16 Q9L1H8	Q9L1H8 streptomyce
36	170.5	14.5	601	5 Q17113	Q17113 babesia bov
37	170	14.4	295	2 Q32850	Q32850 mycobacteri
38	170	14.4	428	10 Q94108	Q94108 zea mays (m
39	170	14.4	548	16 Q06404	Q06404 mycobacteri
40	170	14.4	842	5 Q9VGC9	Q9VGC9 drosophila
41	170	14.4	864	5 Q95U45	Q95U45 drosophila
42	170	14.4	864	5 Q9VGC8	Q9VGC8 drosophila
43	170	14.4	1354	11 Q9EPW8	Q9EPW8 mus musculu
44	169.5	14.4	802	11 P70433	P70433 mus musculu
45	169	14.3	312	5 Q90NNT	Q90NNT drosophila

ALIGNMENTS

RESULT 1	ID	Q9D5J3	PRELIMINARY;	PRT;	344 AA.
AC	Q9D5J3	Q9D5J3	Q9D5J3	Q9D5J3	Q9D5J3
DT	01-JUN-2001	(T-EMBLrel. 17, Created)			
DT	01-JUN-2001	(T-EMBLrel. 17, Last sequence update)			
DT	01-JUN-2001	(T-EMBLrel. 17, Last annotation update)			
DE	4930432K09RIK	protein.			
GN	4930432K09RIK				
OS	Mus musculus (Mouse)				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J; TISSUE=Testis;				
RX	MEDLINE=21085660; PubMed=11217851;				
RA	Arakawa T., Hara A., Fukumitsu Y., Konno H., Adachi Y., Fukuda S.,				
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,				
RA	Saito T., Okazaki Y., Gojoubori T., Bono H., Kasukawa T., Saito R.,				
RA	Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,				
RA	Fleischmann W., Gaasterland T., Glaser C., King B., Kochiwa H.,				
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,				
RA	Schirimi L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,				
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,				
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,				
RA	Brommstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,				
RA	Guernicich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,				
RA	Liyone P., Marchionni L., Mashima J., Mazzarelli J., Monbetsu P.,				
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,				
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,				
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,				
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,				
RA	Hayashizaki Y.;				
RT	Functional annotation of a full-length mouse cDNA collection.;				
RL	Nature 409:685-690(2001).				
DR	EMBL; AK015291; BAB29782.1;				
DR	MGI; MGI:1921029; 4930432K09RIK.				

RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U50901; ABA7806.1; -.
 DR EMBL; AF012905; AAC32526.1; -.
 DR InterPro; IPR000719; Prot_kinase.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 SQ SEQUENCE 1108 AA; 124903 MW; 239CF62D61E11FE7 CRC64;

Query Match 16.0%; Score 188.5; DB 2; Length 1108;
 Best Local Similarity 51.2%; Pred. No. 2.4e-05;
 Matches 42; Conservative 8; Mismatches 27; Indels 5; Gaps 1;

QY 124 PAAPPIAAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAAEPVAAEPAAAP 183
 DB 813 PAAKPEAKPVAAKPEFTTKVAAKPE-----AAKPEAKPVAAKPEAAKPEAAKPE 867

QY 184 VGVPEAAEPSPAEPAATAPAA 205
 DB 868 VAAKPEAAKPVAAKPEAAKPEVA 889

RESULT 13

Q9RX57 PRELIMINARY; PRT; 839 AA.
 AC Q9RX57;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0458.
 GN DR0458.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.U., Lam P., McDonald L., Uitterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001904; AAF10038.1; -.
 DR TIGR; DR0458; -.
 DR InterPro; IPR002965; P-rich extensn.
 DR PRINTS; PR01217; PRICHEXTENS.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 839 AA; 79759 MW; 3B6CC2DCCFEPFADE CRC64;

Query Match 16.0%; Score 188; DB 16; Length 839;
 Best Local Similarity 33.0%; Pred. No. 2e-05;
 Matches 64; Conservative 14; Mismatches 92; Indels 24; Gaps 9;

QY 27 DNDGHPHLSNIPICIRLPPPLYRPVNTPSIPGNTYTTGGLPSYIMLTSP-GFP 85
 DB 140 DPTPAEPLKPP-----VQDTPPPVTPKPTPEPVTPKPAPEPVTAQPPVAKP 193

QY 86 YVYHIGFPLAT-QLNVPLPPLPRGPFVPPSPRFSAAAAPPIAAEPAAAPLTATPV 144
 DB 194 PV-----PAFTSOTPRPPVQAPATRRPPQ---AAFTNNAAGTPAPATQAAPQTP 244

QY 145 AAEPAAAGAPVAAEPAAEPVGAEPAAEPVAAEPAAEPVGAEPAAEPVAAEPAAEP 203
 DB 245 AQAAPATQTPTAPPAAPAAQRPAGAGP---SPAAPAAQNAAPAGSVVPEATVSESSTPA--P 299

QY 204 AAEPHPSPSLQA 217
 DB 300 SAQTP-PTPTRETA 312

RESULT 14

Q9RVY4 PRELIMINARY; PRT; 395 AA.
 AC Q9RVY4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein DR0938.
 GN DR0938.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.U., Lam P., McDonald L., Uitterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001946; AAF10516.1; -.
 DR TIGR; DR0938; -.
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 395 AA; 40003 MW; 75FD963717DAAPFO CRC64;

Query Match 15.8%; Score 186.5; DB 16; Length 395;
 Best Local Similarity 35.0%; Pred. No. 1.2e-05;
 Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSLP-GNTYTDGLSPYMLTSPGFVYHIGFPLATQLNVPLPPLPRGPFVPPSPRP 118
 DB 53 PAEPYGAQPTTSIVSGAVVADSPTQ---AAPAQATAGKTRPA--PAAP----- 102

QY 119 SAAAPPAAP-----IAEPAAAPLTATPVAAEPAAAGPVAAPAAEPVGAEPAA 170
 DB 103 ----APKIPPLPPLPRMPPAPPTPRPVETTTAASPTQPPVTAQATQTPPTQTPAA 158

QY 171 EAPVAAEPAAEPVGAEPAA--EESPAA--EAT-AXGAAP 208
 DB 159 QAPATQPPATPA---EPAAEPAAATTPBPATPEPAAP 198

RESULT 15

Q99KJ1 PRELIMINARY; PRT; 616 AA.
 AC Q99KJ1;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Similar to procollagen, type VI, alpha 3 (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00014; Kunitz_BPTI; 1.
DR Pfam; PF00092; vwa; 1.
DR PRINTS; PRO0059; BASICPTASE.
DR PRINTS; PRO1217; PRICEXTENSN.
DR ProDom; PD000222; Kunitz_BPTI; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00131; KU; 1.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
DR PROSITE; PS02279; BPTI_KUNITZ_2; 1.
DR PROSITE; PS50234; WMPA; 1.
DR Collagen; Protease inhibitor; Serine protease inhibitor.
KW NON_TER
SQ SEQUENCE 1 1 66408 MW; FBA6CAC92B853B3 CRC64;

Query Match 15.8%; Score 186.5; DR 11; Length 616;
Best Local Similarity 32.4%; Pred. No. 1.8e-05;
Matches 67; Conservative 25; Mismatches 72; Indels 43; Gaps 12;

QY 21 PFGEEDNDGHLHPSLNIPIYGIKMLPPPLYRPNVTWSPYSGNTYTDGLPSYPWILT 80
DB 253 PVLAKEDPAKPAQAPAPAKPAKPAKLVPP---QPVHQPA-PAQTASVRPAKP---- 303
QY 81 SPGFPPYVYHIRGFPLATQLNVPELPDPGPFVP---PSRFSSAAAAAPAPPIAAEPAAA 136
DB 304 APPQPA-----AKVPAK--PAVPAQAPAPQPAAPKVPAPKPAVPAQPA 348
QY 137 APLTATPV-----AAEPA-AGAPVAAEPAA--EAPVGAEP-AGAPVAAEPAAEAPV--G 185
DB 349 QPMPAQPVLTGSAAVPAANKFPAKPVATNTATATATAPALAAKPAAPKPAATREPLAAA 408
QY 186 VEPAAEPP-----SPAEPATAKPAAPEP 208
DB 409 VRPVATKPEAPROQAKPAATKPAITTKP 435

Search completed: January 15, 2004, 15:14:11
Job time : 37 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:06:54 ; Search time 18 Seconds

(without alignments)
572.158 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178
Sequence: 1 MKLLIMACIVCAFPARRR.....TAKPAPPEPSPSLQANQ 219

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	180	15.3	555	1	GP1_CHIRE
2	175	14.9	1054	1	IF2_STIAU
3	175	14.9	3149	1	TEGU_EBV
4	174	14.8	518	1	TPM4_DROME
5	170	14.4	3176	1	CA36_HUMAN
6	169.5	14.4	802	1	ENAH_MOUSE
7	169	14.3	299	1	RL22_DROME
8	169	14.3	865	1	CPN_DROME
9	167.5	14.2	450	1	CV1_PARDE
10	166	14.1	810	1	NFM_BOVIN
11	165.5	14.0	439	1	XP2_XENLA
12	162	13.8	5147	1	PCLO_HUMAN
13	160.5	13.6	992	1	EBN6_EBV
14	158	13.4	1274	1	ENAM_MOUSE
15	154	13.1	353	1	CCPA_ACEXY
16	154	13.1	4499	1	DYHA_CHIRE
17	152.5	12.9	5038	1	PCLO_MOUSE
18	152	12.9	721	1	YK82_MYCTU
19	151	12.8	88	1	H82_MISCO
20	151	12.8	211	1	TUBB_SOLTU
21	150	12.7	352	1	ALGP_PSEAE
22	148.5	12.6	464	1	SA32_HUMAN
23	147	12.5	1386	1	ZAP3_MOUSE
24	146.5	12.4	477	1	MAZ_MOUSE
25	145.5	12.4	361	1	IF35_MOUSE
26	145.5	12.4	1565	1	PAC_STRNU
27	145	12.3	475	1	SA32_MOUSE
28	145	12.3	2167	1	SHK1_RAT
29	144.5	12.3	477	1	MAZ_HUMAN
30	142.5	12.1	1083	1	MAZ2_HUMAN
31	142.5	12.1	1083	1	T2D3_HUMAN
32	142	12.1	2161	1	SHK1_HUMAN
33	141.5	12.0	316	1	CDNC_HUMAN

34	141.5	12.0	465	1	FXD1_HUMAN	Q1676	homo sapien
35	140.5	11.9	852	1	WS14_HUMAN	Q9NP71	homo sapien
36	140	11.9	3421	1	TEGU_HAYEB	P28955	equine hiep
37	140	11.9	5085	1	PCLO_RAT	Q9J466	rattus norv
38	139.5	11.8	397	1	SEPL_MOUSE	Q62170	mus musculu
39	139.5	11.8	1157	1	BBC1_YEAST	P47068	saccharomyc
40	138.5	11.8	206	1	FRP1_MEDTR	Q43564	medicago tr
41	138.5	11.8	1183	1	DRPL_RAT	P54258	rattus norv
42	138	11.7	376	1	FXL2_HUMAN	P58012	homo sapien
43	138	11.7	2517	1	NCR2_HUMAN	Q9Y618	h nuclear r
44	137	11.6	236	1	PRP_MEDSA	Q40358	medicago sa
45	136.5	11.6	1003	1	MBDE_HUMAN	Q96dne	homo sapien

ALIGNMENTS

RESULT 1
GPI_CHIRE STANDARD; PRT; 555 AA.
AC Q9FP06; 003927;
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich glycoprotein 1).
GN GPI.
OS Chlamydomonas reinhardtii.
OC Baktaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21159092; PubMed=11258910;
RA Ferris P.J., Moessner J.P., Waffenschmidt S., Kilz S., Drees J.,
RT "Glycosylated polyproline II rods-with-kinks as a structural motif in
RT plant hydroxyproline-rich glycoproteins."
RT Biochemistry 40:2978-2987(2001).
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE FROM N.A.
RX MEDLINE=91017504; PubMed=1699225;
RA Adair W.S., Apt K.B.;
RT "Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
RT encoding cell wall hydroxyproline-rich glycoproteins."
RT Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
CC -1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
CC layer.
CC -1- SUBUNIT: Associates with GP2 and GP3.
CC -1- PTM: N-glycosylated and O-glycosylated.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.sib.ch).
CC -----
CC EMBL; AF309494; AAC45420.1; -;
DR EMBL; M58496; AAA69706.1; ALT_SEQ.
DR GlycosubitedB; Q9FP06; -;
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR003882; Pfam1_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR01218; PSTEXTENSIN.
DR Glycoprotein; Repeat; Signal.
KW SIGNAL.
FT CHAIN 30 555
FT DOMAIN 40 339
FT DOMAIN 259 279
FT CARBOHYD 399 399
FT CARBOHYD 455 455
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455

[illegible]

ID	IP2_STIAU	STANDARD;	PRT;	1054 AA.
AC	P55875;			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	translation	initiation factor IP-2.		
GN	INPB.			
OS	Stigmatella aurantiaca.			
OC	Bacteria; Proteobacteria; Delaproteobacteria; Myxococcales;			
OC	Cyctobacteriineae; Cyctobacteraceae; Stigmatella.			
OX	NCBI_TaxID=41;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DM4;			
RX	MEDLINE=97234648; PubMed=9079922;			
RA	Bremond L., Laalami S., Derijard B., Cernatempo Y.;			
RT	"translation initiation factor IP2 of the myxobacterium Stigmatella			
RT	aurantiaca: presence of a single species with an unusual N-terminal			
RT	sequence.";			
RL	J. Bacteriol. 179:2348-2355 (1997).			
CC	-1- FUNCTION: One of the essential components for the initiation of			
CC	protein synthesis. Protects formylmethionyl-tRNA from spontaneous			
CC	hydrolysis and promotes its binding to the 30S ribosomal subunits.			
CC	Also involved in the hydrolysis of GTP during the formation of the			
CC	70S ribosomal complex (By similarity).			
CC	-1- SUBCELLULAR LOCATION: Cytoplasmic.			
CC	-1- SIMILARITY: BELONGS TO THE IP-2 FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration of			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X87940; CAA61162.1; -			
DR	PIR; T43226; T43226.			
DR	HSSP; P13551; IELO.			
DR	HAWAP; MF_00100; -; 1.			
DR	InterPro; IPR000795; EF_GTPbind.			
DR	InterPro; IPR004161; EFTU_D2.			
DR	InterPro; IPR000178; IP2.			
DR	InterPro; IPR006847; IP2_N.			
DR	InterPro; IPR005225; Small_GTP.			
DR	Pfam; PF00009; GTP_EFTU; 1.			
DR	Pfam; PF01144; GTP_EFTU_D2; 2.			

DR	Pfam; PF04760; IP2_N; 2.	
DR	ProDom; PD186100; IP2; 1.	
DR	TIGRFAMs; TIGR00487; IP-2; 1.	
DR	TIGRFAMs; TIGR00231; small_GTP; 1.	
DR	PROSITE; PS01176; IP2; 1.	
KW	Initiation factor; Protein biosynthesis; GTP-binding.	
FT	DOMAIN	555 703
FT	NP_BIND	561 568
FT	NP_BIND	607 611
FT	NP_BIND	661 664
SO	SEQUENCE	1054 AA; 111323 MW; FFD519530B5D0669 CRC64;

```

Query Match      14.9%; Score 175; DB 1; Length 1054;
Best Local Similarity 38.3%; Pred. No. 0.0029;
Matches 49; Conservative 8; Mismatches 51; Indels 20; Gaps 4

QY    96 ATQTNLVPELPGRGPPFVYPSPRRFSAAAAAPAAPPIAA--EPAAAABELTAT--PVAAPPAAG 151
       : :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
DB    98 ASDVSPPSPD-----VHEASGAEMAAASERVFBAAAAGEPVAAAPRAAASEBPAAPKAT 152

QY    152 APVAAEAPAAAEFVGAEAPAAAEAPVAAAEAPAAAEAPGVGEAPAAEEBSP-----AEPRT 200
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB    153 APVAPBEPTVBAPKAAPAPVAPBEPTVEAPKTEAPVAAAAPTAEAPTBPARTETVPVTSGRRAAS 212
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY    201 AKPAAPRP 208
       : : : : :
DB    213 CRGAAPLC 220

```

```

RESULT 3
TEGU EBV          STANDARD;          PRT;    3149 AA.
AC      P03186;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-JUL-1986 (Rel. 01, Last sequence update)
DT      01-APR-1993 (Rel. 25, Last annotation update)
DE      Large tegument protein.
CN      Bp.F1.
OC      Epstein-Barr virus (strain B95-8) (Human herpesvirus 4) .
OC      Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC      Gammaherpesvirinae; Lymphocryptovirus.
OX      NCBI_TaxID=10377;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=84270667; PubMed=6087149;
RA      Baar R., Bankier A.T., Blagijn M.D., Deininger P.L., Farrell P.J.,
RA      Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA      Tufnell P.S., Watfrell B.G.;
RT      "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL      Nature 310:207-211(1984) .
CC      -1- FUNCTION: TEGUMENT PROTEIN.
CC      -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
CC      HSV-1 24, EBV BpLF1, HSV-1 64, VZV 22, AND HCMV UL46.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.1db-6ib.ch/announce/
CC      or send an email to license@1db-6ib.ch).
CC      -----
CC      EMBL; V01555; CAA24839.1; -.
DR      PIR; G93065; Q0858.
DR      InterPro; IPR005928; Herpes_teg_N.
DR      Pfam; PF08483; Herpes_teg_N.1.
SQ      SEQUENCE      3149 AA; 337954 MW; 3DD0C576587313D8 CRC64;

Query March          14.9%; Score 175; DB 1; Length 3149;
Best Local Similarity 34.0%; Pred. No. 0.0071;
Matches 48; Conservative 19; Mismatches 40; Indels 34; Gaps 6;

105 PRRGPP-----VP-----DSRFSSAAAPPA-----PPIAEPAAAPLTATPA 145

```

```

Db      290 PPVWLPFRKRIITPVRLRPLPSISFTSDSPFAARKYSPAKTNSPP--SSPASAAPASAPAS 347
Qy      146 AAPAAGAEVAAEPAAAEPAVGAEPPAAEAPVVAEPAAAEPAVGVPEPAAEPEEPAPATAKP-A 204
Db      348 AAPASAAAPASAAAPASAAAPASAAAPASAPSPPLPIPIGLDHTGVPAPSPTRPPASSGA 407
Qy      205 APE-----PHSPS 213
Db      408 APQTPRKRGKGLKQDSPHKKPT 428

RESULT 4
TPM4_DROME STANDARD; PRT; 518 AA.
ID TPM4_DROME STANDARD; PRT; 518 AA.
AC P49455; P49456; Q24425; Q24426;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Tripomysosin 1, Isoforms 33/34 (Tripomysosin II).
GN TM1 OR TM11.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota, Metazoa, Arthropoda, Hexapoda, Insecta, Pterygota;
OC Neoptera, Endopterygota, Diptera, Brachycera, Muscomorpha;
OC Ephydroidea, Drosophilidae, Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC TISSUE=Embryo, and Pupa;
RX MEDLINE=89127197; PubMed=2851721;
RA Hanke P.D., Scott R.V.;
RT "The Drosophila melanogaster tropomyosin II gene produces multiple
RT proteins by use of alternative tissue-specific promoters and
RT alternative splicing."
RT Mol. Cell. Biol. 8:3591-3602(1988).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 33 AND 34).
RC STRAIN=Oregon-R; TISSUE=Pupa;
RX MEDLINE=87064486; PubMed=3097506;
RA Karlik C.C., Fyrberg E.A.;
RT "Two Drosophila melanogaster tropomyosin genes: structural and
RT functional aspects."
RT Mol. Cell. Biol. 6:1965-1973(1986).
RN [3]
RP -1- ALTERNATIVE PRODUCTS:
RC Event=Alternative splicing; Named isoforms=5;
RX Comment=Additional isoforms seem to exist;
CC Name=33; Synonyms=9C;
CC Name=Isocle; Synonyms=9D;
CC Name=Isocle; Synonyms=9D;
CC IsoId=P06754-1; Sequence=External;
CC Name=Non-muscle; Synonyms=Cytoskeletal;
CC IsoId=P06754-2; Sequence=External;
CC Name=9A;
CC IsoId=P06754-3; Sequence=External;
CC Name=34; Synonyms=9B;
CC IsoId=P49455-2; Sequence=VSP 006623, VSP 006624, VSP 006625;
CC -1- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect
CC flight muscles
CC -1- DEVELOPMENTAL STAGE: Both isoforms are expressed during pupal and
CC adult stages.
CC -1- DOMAIN: THE MOLECULE IS IN A COILED COIL STRUCTURE. THE SEQUENCE
CC EXHIBITS A PROMINENT SEVEN-RESIDUES PERIODICITY.
CC -1- SIMILARITY: BELONGS TO THE TROPOMYOSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; X76208; CAA53800.1; -.

```

[illegible]

Db 504 ABAAPAAABGEAPPA 518

RESULT 5
CA36_HUMAN STANDARD; PRT; 3176 AA.
ID CA36_HUMAN
AC P12111; Q16501;
DT 01-OCT-1989 (Rel. 12, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 3 (VI) chain precursor.
GN COL6A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NC NCB1; TaxID=9606;
RX MEDLINE=21987636; PubMed=11992252;
RC TISSUE=Fibroblast;
RA Chu M.-L., Zhang R.-Z., Pan T.-C., Stokes D., Conway D., Kuo H.-J.,
RA Glanville R., Mayer U., Mann K., Deutzmann R., Timpl R.;
RT Mosaic structure of globular domains in the human type VI collagen
RT alpha 3 chain: similarity to von Willebrand factor, fibronectin,
RT actin, salivary proteins and apocrotinin type protease inhibitors.";
RL EMBO J. 9:385-393(1990).
RN [2]
RP REVISIONS.
RA Chu M.-L.;
RN Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
RP SEQUENCE OF 2038-2373 FROM N.A.
RX MEDLINE=89066644; PubMed=3198591;
RA Chu M.-L., Conway D., Pan T.-C., Baldwin C., Mann K., Deutzmann R.,
RA Timpl R.;
RT "Amino acid sequence of the triple-helical domain of human collagen
RT type VI.";
RL J. Biol. Chem. 263:18601-18606(1988).
RN [4]
RP SEQUENCE OF 2092-2157 FROM N.A.
RX MEDLINE=88029444; PubMed=3665927;
RA Chu M.-L., Mann K., Deutzmann R., Pribula-Conway D.,
RA Hu-Chen C.-C., Bernard M.P., Timpl R.;
RT "Characterization of three constituent chains of collagen type VI by
RT peptide sequences and cDNA clones.";
RL Eur. J. Biochem. 168:309-317(1987).
RN [5]
RP SEQUENCE OF 2092-2151 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=88161046; PubMed=3348213;
RA Weil D., Martel M.-G., Passage E., van Cong N., Pribula-Conway D.,
RA Mann K., Deutzmann R., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal localization of human genes encoding the
RT three chains of type VI collagen.";
RL Am. J. Hum. Genet. 42:435-445(1988).
RN [6]
RP SEQUENCE OF 32-236 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=93054780; PubMed=1339440;
RA Zanussi S., Doliana R., Segat D., Bonaldo P., Colombatti A.;
RT "The human type VI collagen gene. mRNA and protein variants of the
RT alpha 3 chain generated by alternative splicing of an additional 5-end
RT exon.";
RL J. Biol. Chem. 267:24082-24089(1992).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 3107-3164.
RX MEDLINE=95182468; PubMed=7533217;
RA Arnoux B., Merigean K., Saludjian P., Norris F., Norris K., Bjoern S.,
RA Olsen O., Petersen L., Ducruix A.;
RT "The 1.6 A structure of Kunitz-type domain from the alpha 3 chain of
RT human type VI collagen.";
RL J. Mol. Biol. 246:609-617(1995).
RN [8]

RP STRUCTURE BY NMR OF 3102-3164.
RX MEDLINE=96398604; PubMed=8805527;
RA Zweckstetter M., Czisch M., Mayer U., Chu M.-L., Zinth W., Timpl R.,
RA Holak T.A.;
RT "Structure and multiple conformations of the Kunitz-type domain from
RT human type VI collagen alpha3 (VI) chain in solution.";
RL Structure 4:195-209(1996).
RN [9]
RP STRUCTURE BY NMR OF 3107-3164.
RX MEDLINE=97410331; PubMed=9256624;
RA Soerensen M.D., Bjoern S., Norris K., Olsen O., Petersen L.,
RA James T.L., Led J.J.;
RT "Solution structure and backbone dynamics of the human alpha3-chain
RT type VI collagen C-terminal Kunitz domain.";
RL Biochemistry 36:10439-10450(1997).
RN [10]
RP DISEASE.
RX MEDLINE=21987636; PubMed=11992252;
RA Demir E., Sabatelli P., Allamand V., Ferreira A., Moghadasezadeh B.,
RA Makrelouf M., Topaloglu H., Echeine B., Merini L., Guichenev P.;
RT "Mutations in COL6A3 cause severe and mild phenotypes of Ulrich
RT congenital muscular dystrophy.";
RL Am. J. Hum. Genet. 70:1446-1458(2002).
RN [11]
RP VARIANT BM GLU-1679, AND VARIANT HIS-2831.
RX MEDLINE=98204804; PubMed=9536084;
RA Pan T.-C., Zhang R.-Z., Pericak-Vance M.A., Tandan R., Fries T.,
RA Stajich J.W., Viles K., Vance J.M., Chu M.-L., Spear M.C.;
RT "Missense mutation in a von Willebrand factor type A domain of the
RT alpha 3 (VI) collagen gene (COL6A3) in a family with Bethlem
RT myopathy.";
RL Hum. Mol. Genet. 7:807-812(1998).
CC -1- FUNCTION: COLLAGEN VI ACTS AS A CELL-BINDING PROTEIN.
CC -1- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1 (VI),
CC -1- ALPHA 2 (VI), AND ALPHA 3 (VI).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named Isoforms=2;
CC Name=1;
CC Name=2;
CC Name=1; IsoId=P12111-1; Sequence=Displayed;
CC Name=2; IsoId=P12111-2; Sequence=VSP 001172;
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- DISEASE: Defects in COL6A3 are a cause of Bethlem myopathy (BM)
CC [MIM:158810]. BM is a rare autosomal dominant proximal myopathy
CC characterized by early childhood onset (complete penetrance by the
CC age of 5) and joint contractures most frequently affecting the
CC elbows and ankles.
CC -1- DISEASE: Defects in COL6A3 are a cause of Ulrich congenital
CC muscular dystrophy (UCMD) [MIM:254090]; also known as Ulrich
CC scleroatonic muscular dystrophy, an autosomal recessive congenital
CC myopathy. UCMD is characterized by muscle weakness and multiple
CC joint contractures, generally noted at birth or early infancy. The
CC clinical course is more severe than in Bethlem myopathy.
CC -1- SIMILARITY: Contains 1 BPPI/Kunitz inhibitor domain.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- SIMILARITY: Contains 12 VMPA domains.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; X52022; CAA36267.1; -;
CC EMBL; X06196; CAA29557.1; -;
CC EMBL; M20778; -; NOT ANNOTATED_CDS.
CC EMBL; M27449; AAB52057.1; -;
CC EMBL; S69432; AAB34261.1; -;
CC EMBL; A59140; CGB03A.
CC PIR; AS9140; CGB03A.
CC PDB; 1KNT; 01-NOV-94.

DR PDB; 2KNT; 15-MAY-97.
 DR PDB; 1KUN; 12-NOV-97.
 DR PDB; 1KTH; 28-AUG-02.
 DR Genew; HGNC:2213; COL6A3.
 DR MIM; 120250; -.
 DR MIM; 158810; -.
 DR MIM; 254050; -.
 DR GO; GO:0005589; C:collagen type VI; TAS.
 DR GO; GO:0005202; F:collagen; TAS.
 DR GO; GO:0007517; P:muscle development; TAS.
 DR InterPro; IPR00087; Collagen.
 DR InterPro; IPR00361; FN_III.
 DR InterPro; IPR002223; Kunitz_BPTI.
 DR InterPro; IPR02035; WVF_A.
 DR Pfam; PF001391; Collagen_5.
 DR Pfam; PF00014; Kunitz_BPTI; 1.
 DR Pfam; PF00092; WVF; 11.
 DR PRINTS; PR00759; BASICPRASE.
 DR ProDom; PD000007; C1g_helix; 2.
 DR ProDom; PD000222; Kunitz_BPTI; 1.
 DR SMART; SM00060; FN3; 1.
 DR SMART; SM00131; KU; 1.
 DR SMART; SM00327; WVF; 12.
 DR PROSITE; PS00280; BPTI_KUNITZ_1; 1.
 DR PROSITE; PS50279; BPTI_KUNITZ_2; 1.
 DR PROSITE; PS50234; WVF; 12.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Cell adhesion; Serine protease inhibitor; Collagen;
 KW Signal; 3D-structure; Disease mutation; Polymorphism;
 KW Alternative splicing.
 KM
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 3176 COLLAGEN ALPHA 3 (VI) CHAIN.
 FT DOMAIN 26 2038 NONHELICAL REGION.
 FT DOMAIN 2039 2375 TRIPLE-HELICAL REGION.
 FT DOMAIN 2376 3176 NONHELICAL REGION.
 FT DOMAIN 39 213 WVF_A 1.
 FT DOMAIN 242 419 WVF_A 2.
 FT DOMAIN 445 620 WVF_A 3.
 FT DOMAIN 639 816 WVF_A 4.
 FT DOMAIN 837 1009 WVF_A 5.
 FT DOMAIN 1029 1205 WVF_A 6.
 FT DOMAIN 1233 1404 WVF_A 7.
 FT DOMAIN 1436 1609 WVF_A 8.
 FT DOMAIN 1639 1812 WVF_A 9.
 FT DOMAIN 1838 2024 WVF_A 10.
 FT DOMAIN 2402 2581 WVF_A 11.
 FT DOMAIN 2619 2815 WVF_A 12.
 FT DOMAIN 2887 3076 FIBRONECTIN TYPE-III.
 FT DOMAIN 3107 3176 BPTI/KUNITZ INHIBITOR.
 FT SITE 2040 2042 CELL ATTACHMENT SITE.
 FT SITE 2136 2138 CELL ATTACHMENT SITE.
 FT SITE 2148 2150 CELL ATTACHMENT SITE.
 FT SITE 2154 2156 CELL ATTACHMENT SITE.
 FT SITE 2370 2372 CELL ATTACHMENT SITE.
 FT ACT_SITE 3121 3122 REACTIVE BOND.
 FT DISULFID 3111 3161
 FT DISULFID 3120 3144
 FT DISULFID 3136 3157
 FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 251 251 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT
 Query Match 14.4%; Score 170; DB 1; Length 3176;
 Best Local Similarity 31.8%; Pred. No. 0.014;
 Matches 57; Conservative 16; Mismatches 56; Indels 50; Gaps 9;

Qy 151 GAPVAAEPAAEA-DVGAEPAAEAFAVGVGPAAPAEPSAPAPATKAPAP 208
 Db 2933 --PVAIVPATAKAVAAKPAAPVPPA---AAKAVATKPEVPRPQAPATKATKTP 2986
 RESULT 6
 ENAH_MOUSE STANDARD; PRT; 802 AA.
 AC Q03173; P70430; P70431; P70432; P70433;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Enabled protein homolog (NPC derived proline-rich protein 1) (NDPP-1).
 GN ENAH OR MENA OR NDPEL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain;
 RX MEDLINE=93041923; PubMed=1420303;
 RA Sazuka T., Tomooka Y., Kachju S., Ikawa Y., Noda M., Kumar S.;
 RT "Identification of a developmentally regulated gene in the mouse
 central nervous system which encodes a novel proline rich protein.";
 RL Biochim. Biophys. Acta 1132:240-248(1992).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4 AND 5), FUNCTION, AND SUBCELLULAR
 RP LOCATION.
 RC TISSUE=Brain;
 RX MEDLINE=97015079; PubMed=8861907;
 RA Gertler F.B., Niebur K., Reinhard M., Wehland J., Soriano P.;
 RT "Mena, a relative of VASP and Drosophila Enabled, is implicated in the
 control of microfilament dynamics.";
 RL Cell 87:227-239(1996).
 RN [3]
 RP FUNCTION, AND SUBUNIT.
 RX MEDLINE=99166867; PubMed=1069337;
 RA Lanier L.M., Gates M.A., Witke W., Menzies A.S., Wehman A.M.,
 RA Macklis J.D., Kwiatkowski D., Soriano P., Gertler F.B.;
 RT "Mena is required for neurulation and commissure formation.";
 RL Neuron 22:313-325 (1999).
 CC -1- FUNCTION: May be involved in microfilament assembly and cell
 motility. Induces the formation of F-actin rich outgrowths in
 fibroblasts. Required for neurulation and commissure formation.
 CC -1- SUBUNIT: Binds profilin.
 CC -1- SUBCELLULAR LOCATION: Localized to focal adhesions and, to a
 lesser extent, leading edges and stress fibers.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=5;
 CC Name=5; Synonym=Mena+++;
 CC IsoId=Q03173-1; Sequence=Displayed;
 CC Name=1;
 CC IsoId=Q03173-2; Sequence=VSP_007255, VSP_007256;
 CC Name=2; Synonym=Mena;
 CC IsoId=Q03173-3; Sequence=VSP_007259, VSP_007260;
 CC Name=3; Synonym=Mena+;
 CC IsoId=Q03173-4; Sequence=VSP_007259;
 CC Name=4; Synonym=Mena++;
 CC IsoId=Q03173-5; Sequence=VSP_007257, VSP_007258;
 CC -1- TISSUE SPECIFICITY: IN THE HEART AND TESTIS AND LESS SO IN THE
 CC LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS AND BRAIN.
 CC -1- SIMILARITY: Contains 1 WH1 domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1db-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; D10727; BAA01570.1; --


```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - -
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation - -
CC CC the European Bioinformatics Institute. There are no restrictions on its use. By non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.ebi.ac.uk/announcement/
CC CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, M17522; AAA25573.1; -.
DR EMBL, X05799; CAA29245.1; -.
DR PIR, C29413; C29413.
DR InterPro, IPR002326; CyC_C1.
DR InterPro, IPR000345; CyTC_heme_bind.
DR Pfam, PF02167; Cytochrome_C1_1.
DR PRINTS, PR00603; CYTOCHROME_C1.
DR PROSITE, PS00190; CYTOCHROME_C_1.
DR Electron transport; Respiratory chain; Heme; Transmembrane, Signal.
KM SIGNAL 1 21
FT CHAIN 1 22 450 CYTOCHROME C1.
FT BINDING 245 245 HEME (COVALENT).
FT BINDING 248 248 HEME (COVALENT).
FT METAL 249 249 IRON (HEME AXIAL LIGAND).
FT METAL 373 373 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT TRANSMEM 421 435 ANCHORS TO THE MEMBRANE (POTENTIAL).
SQ SEQUENCE 450 AA; 46874 MW; CLEA8B6E48FE7603 CRC64;

Query Match 14.2%; Score 167.5; DB 1; Length 450;
Beat Local Similarity 42.6%; Pred. NO. 0.0037;
Matches 58; Conservative 8; Mismatches 35; Indels 35; Gaps 8;

OY 114 PSRRFSAAAAPAPPIAAEPAAA--APLTATPVA-----EPAGAPVA-----AEPA 159
Db 48 PAADDTAAEAEMADBPAAEEAAGRAVETEEBPAAATETPAEPADDEPAAETEPAAEEPA 107
OY 160 -----AEPVGAEPAAEPVAAEPAAEPVGEVPEAAEESPAAEP--ATAPAA 205
Db 108 AEEAQNTTEEAEPAAEPAAEPAAEP--AEPAAADP--AEAAAEEAEPAAEPAAEPAA 164
OY 206 PEP---HPSPSLREQA 217
Db 165 EEPETEETEEAEPAAEENA 180

RESULT 10
NFM_BOVIN STANDARD; PRT; 810 AA.
ID NFM_BOVIN
AC 077788;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DS (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN NEF3 OR NEFM OR NFM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OC NCBI_TaxId=9913;
OK (1)
RP SEQUENCE FROM N.A.
RA Hill W.D., Zhang L., Ballin B.J., Sprinkle T.J., Spicer K.,
RA Gearhart D.A.;
RT "The bovine neurofilament M subunit has a novel set of KSP repeats
RT normally restricted to NF-H."
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -1- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLET K-S-P. NFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER (BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING

```

```

CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC CC or send an email to license@ebi.ac.uk).
CC CC -----
CC DR EMBL; AF091342; AAC36357.1; -.
CC DR InterPro; IPR001664; IP.
CC DR InterPro; IPR002957; Keratin_1.
CC DR Pfam; PF00038; filament; 1.
CC DR PRINTS; PRO1248; TYPE1KERATIN.
CC DR PROSITE; PS00226; IF, 1.
CC KW Intermediate filament; Coiled coil; Neutrone; Phosphorylation.
CC FT NON TRR 1
CC FT DOMAIN <1 296 ROD.
CC FT DOMAIN 297 810 TAIL.
CC FT DOMAIN <1 30 COIL 1A.
CC FT DOMAIN 21 33 LINKER 1.
CC FT DOMAIN 34 132 COIL 1B.
CC FT DOMAIN 133 149 LINKER 12.
CC FT DOMAIN 150 171 COIL 2A.
CC FT DOMAIN 172 175 LINKER 2.
CC FT DOMAIN 176 296 COIL 2B.
CC FT DOMAIN 503 582 8 X 13 AA TANDEM REPEATS.
CC ST SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;
CC SQ
CC
CC Query March 14.1%; Score 166; DB 1; Length 810;
CC Best Local Similarity 35.1%; Pred. No. 0.0073;
CC Matches 34; Conservative 12; Mismatches 51; Indels 0; Gaps 0;
CC
CC QY 112 VPSRPFSSAAAPAPPIAPBAAAADLTATPVAAEPAAGAPVAAPAEAPVGAEPAAE 171
CC DB 484 VAPGEELAAAKVKEPKAKSPVAKSPPTKSPPTAKSPKAKSPPTAKSPVAK 543
CC QY 172 APVAAPAAAPVGVPEAAEPPAPAPATKAPAPAP 208
CC DB 544 SPTAKSPKAKSPKAKSPPTAKSPPTAKSPKAKSP 580
CC
CC RESULT 11
CC XP2_XENLA STANDARD; PRT; 439 AA.
AC ID_XP2_XENLA STANDARD; PRT; 439 AA.
AC P17437; Q08944;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DT Skin secretory protein xp2 precursor (ABEG protein).
OS Xenopus laevis (African Clawed Frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
RX MEDLINE=92332564; PubMed=1628230.
RA Hauser F., Roeben C., Hoffmann W.;
RT "xp2, a new member of the P-domain peptide family of potential growth
RT factors, is synthesized in Xenopus laevis skin.";
RL J. Biol. Chem. 267:14451-14455(1992).
RN [2]
RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
RC TISSUE=Skin.
RX MEDLINE=90127399; PubMed=2298293.
RA Gmachl M., Berger H., Thalhammer J., Kreil G.;
RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly

```

RT repetitive amino acid sequence."
 RL PDB Lett. 260:145-148(1990).
 CC -1- FUNCTION: MAY ACT AS A GROWTH FACTOR IN THE GERMINAL LAYER OF THE
 CC EPIDERMIS. MAY ALSO BE INVOLVED IN GROWTH OF REGENERATING GLANDS
 CC AND IN PROTECTION OF THE SKIN FROM THE EXTERNAL ENVIRONMENT.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1; Synonyms=ABEG;
 CC IsoId=P17437-1; Sequence=Displayed;
 CC Name=2; Synonyms=XP2;
 CC IsoId=P17437-2; Sequence=VSP_004652;
 CC -1- TISSUE SPECIFICITY: SKIN.
 CC -1- SIMILARITY: Contains 2 P-type (Pcfo1) domains.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 392
 CC ONWARD AND IS SHORTER (418 AA) DUE TO A FRAMESHIFT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M90095; AAA50001.1; -;
 DR EMBL; X51394; CA35759.1; ALT_FRAME.
 DR PIR; A37331; A37331.
 DR HSSP; P04155; 1P82.
 DR InterPro; IPR000519; P_cfo1.
 DR Pfam; PF00088; trefol1.2.
 DR PRINTS; PR00680; PTFREFOL1.
 DR SMART; SM00018; PD; 2.
 DR PROSITE; PS00025; P_TREFOL1; 2.
 DR Signal; Growth factor; Alternative splicing; Repeat;
 KM Signal; Growth factor; Alternative splicing; Repeat;
 FT STGNAL 1 22 POTENTIAL.
 FT CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
 FT MOD RES 23 23 PYROLIDONE CARBOXYLIC ACID (PROBABLE).
 FT DOMAIN 26 343 33 X REPEATS OF G-[EB]-[AP] (2,4)-A-E.
 FT DOMAIN 350 391 P-TYPE 1.
 FT DOMAIN 397 438 P-TYPE 2.
 FT DISULFID 351 377 BY SIMILARITY.
 FT DISULFID 371 376 BY SIMILARITY.
 FT DISULFID 371 388 BY SIMILARITY.
 FT DISULFID 398 424 BY SIMILARITY.
 FT DISULFID 408 423 BY SIMILARITY.
 FT DISULFID 418 435 BY SIMILARITY.
 FT VARSPLIC 26 343 Missing (in isoform 2).
 FT CONFLICT 3 3 H -> S (IN REF. 2).
 FT CONFLICT 18 18 C -> W (IN REF. 2).
 SQ SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 14.0%; Score 165.5; DB 1; Length 439;
 Best Local Similarity 39.7%; Pred. No. 0.0047;
 Matches 48; Conservative 6; Mismatches 56; Indels 11; Gaps 3;

QY 103 PLPPRGPPFPSPSPFSAAAAAPPAAPPAAPPAAPPAAPPAAPPAAG-APVAAPPAE 161
 DB 103 PAPAEGAPAP-----AEGGAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAE 157
 QY 162 APVGAEPAPAPVAPAPAPAPVGVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAE 216
 DB 158 APAE 217
 QY 217 A 217
 DB 218 A 218

RESULT 12
 PCIO_HUMAN

ID PCIO_HUMAN STANDARD; PRT; 5147 AA.
 AC Q9Y6V0; Q43373; O60305; Q9BVC8; Q9UIV2; Q9Y6U9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Piccolo protein (Aczonin) (Fragments).
 OS PCIO OR ACZ OR KIA0559.
 GN Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE OF 1-759 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99439764; PubMed=10508862;
 RA Wang X., Kibschull M., Lane M.M., Lichte B., Petrasch-Parwez E.,
 RA Kiliann M.W.,
 RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
 RT zones, shares homology regions with rlm and bassoon and binds
 RT profilin." J. Cell Biol. 147:151-162(1999).
 RL [2]
 RN SEQUENCE OF 552-4404 FROM N.A.
 RP Kraemer J., Wollam C., Wohldmann P., McGrane B.;
 RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 3619-5147 FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=9829545; PubMed=9628581;
 RA Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
 RA Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. IX.
 RT The complete sequences of 100 new cDNA clones from brain which can
 RT code for large proteins in vitro." DNA Res. 5:31-39(1998).
 RL [4]
 RN SEQUENCE OF 4405-4439 FROM N.A.
 RP TISSUE=Placenta;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Peigold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan K., Moore T., Max S.I., Wang J., Helel F.,
 RA Diatchenko L., Marusik K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ueda T.B., Tohyuki S., Carninci P., Prange C.,
 RA Rana S.S., Loquellano N.A., Peters G.D., Abramson R.D., Mulany S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Munz D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kertman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Schenker A., Schein J.B., Jones S.J.M., Maria M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences." Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RL [5]
 RN SEQUENCE OF 4405-5147 FROM N.A.
 RP Kallio J., Elliott G.;
 RA Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: May act as a scaffolding protein involved in the
 CC organization of synaptic active zones and in synaptic vesicle
 CC trafficking (By similarity).
 CC -1- SUBUNIT: Interacts with Rabac1/Prai and profilin (By similarity).
 CC -1- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
 CC synaptic junctions (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Comment=Additional isoforms seem to exist;
 CC Name=1;

CC IsoId=Q9Y6V0-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9Y6V0-2; Sequence=VSP_003923, VSP_003924, VSP_003925,
 CC VSP_003926, VSP_003927;
 CC Note=No experimental confirmation available;
 CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
 CC phospholipids. Calcium binds with low affinity but with high
 CC specificity and induces a large conformational change.
 CC -1- SIMILARITY: Contains 2 C2 domains.
 CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Y19188; CAB60727.1; -;
 CC EMBL; AC004903; AAD20936.1; -;
 CC EMBL; AC004886; AAD21789.1; -;
 CC EMBL; AB011331; BAA25485.1; -;
 CC EMBL; BC001304; AAH01304.1; -;
 CC EMBL; AC004082; AAB97937.1; -;
 CC PIR; T00634; T00634.
 CC HSSP; P04410; 1A25.
 CC Genew; HGNC:13406; PCLO.
 CC MIM; 604918; -;
 CC GO; GO:0005856; C:cytoskeleton; NAS.
 CC GO; GO:0045202; C:synaptic junction; ISS.
 CC GO; GO:0005509; F:calcium ion binding activity; ISS.
 CC GO; GO:0005544; F:calcium-dependent phospholipid binding acti. .; ISS.
 CC GO; GO:0005522; F:profilin binding activity; ISS.
 CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
 CC GO; GO:0016080; P:synaptic vesicle targeting; ISS.
 CC InterPro; IPR000008; C2.
 CC InterPro; IPR001565; Synaptoctagmin.
 CC PRINTS; PRO0360; C2DOMAIN.
 CC PRINTS; PRO0399; SYNAPTOTAGMIN.
 CC SMART; SM00239; C2; 2.
 CC PROSITE; PS00499; C2 DOMAIN 1; 1.
 CC PROSITE; PS00004; C2 DOMAIN 2; 2.
 CC KMW; Calcium/phospholipid-binding; zinc; Metal-binding; zinc-finger;
 CC Repeat; Alternative splicing.
 CC FT NON TER 1 1
 CC FT 400 465
 CC FT 499 523
 CC FT 969 992
 CC FT 1010 1011
 CC FT 2300 2325
 CC FT 4391 4442
 CC FT 4544 4633
 CC FT 5031 5121
 CC FT 4404 4404
 CC FT 4534 4534
 CC FT 4576 4576
 CC FT 4757 4761
 CC FT 5147 5147
 CC FT 563537 MM; CDSB64990496CDDC CRC64;
 CC SEQUENCE 5147 AA; 563537 MM; CDSB64990496CDDC CRC64;
 CC Query Match 13.8%; Score 162; DB 1; Length 5147;
 CC Best Local Similarity 23.9%; Pred. No. 0.057;
 CC Matches 60; Conservative 31; Mismatches 78; Indels 82; Gaps 10;
 CC 31 GHFHLPSLNIPIYGRILPPPLVYRPVNTVPSYGNNTYTDG---LSSYPMILTSPPFPV 87

DB 255 GESVYPSLPSF-----SKPPIQPTFGKPPAQPGHRSQPPAPKPPAQPSGLTK----- 304
 QY YHNGFPLATQANV--PPLPRGF-----PVPSPRFFSA----- 120
 DB 305 -----PLAQPGGVYKPPVQPPGTTKPPAQPLGAKKPPAQQTGSEKSSSQPGPKALAQ 358
 QY 121 -----AAAPAP-----PIAEPAAAPL-----TATPVAEPAPAPVAE 157
 DB 359 PGVGTTPAQGGPAKPPVQGVGTPEKPLAQGQLGSPAPAPPTKTPAQTKPPSQPGSGTK 418
 QY 158 PAAPAPVAEPAPAPAPVAEPAPAPVGEPAEPSPAPAT-----AKPA 205
 DB 419 PPPQPPGPAKSPPOPGSTKPPSQPGSAKSPAQPPAPKSAQQTFRKPSQTGFGKPLQ 478
 QY 206 PEPHPSPLEQ 216
 DB 479 P-PTVSPSAKQ 488
 RESULT 13
 EBN6_EBV STANDARD; PRT; 992 AA.
 ID EBN6_EBV
 AC P03204;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE EBN6-6 nuclear protein (EBNA-3C) (EBNA-4B).
 GN BBRP3-BBRP4.
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10377;
 RN [1]
 RP SOURCE: FROM N.A.
 RX MEDLINE=84270667; PubMed=6087149;
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
 RA Gibson T., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
 RA Tufnell P.S., Watnall B.G.;
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
 RL Nature 310:207-211 (1984).
 RN [2]
 RP CHARACTERIZATION.
 RX MEDLINE=88155772; PubMed=2831394;
 RA Petri L., Sample J., Wang F., Kleff B.;
 RT "A fifth Epstein-Barr virus nuclear protein (EBNA3C) is expressed in
 RT latently infected growth-transformed lymphocytes.";
 RL J. Virol. 62:1330-1338 (1988).
 RN [3]
 RP SUBCELLULAR LOCATION.
 RX MEDLINE=90266473; PubMed=2161150;
 RA Petri L., Sample C., Kleff B.;
 RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
 RT latent infection nuclear proteins.";
 RL Virology 176:563-574 (1990).
 CC -1- FUNCTION: INVOLVED IN LATENT CYCLE.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
 CC -1- SIMILARITY: SOME SIMILARITIES EXIST BETWEEN EBNA 4, 5, AND 6.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; V01555; CAA24859.1; -;
 CC Pfam; PF05009; EBNA-3; 1.
 CC Nuclear protein; Repeat.
 CC FT DOMAIN 74 80
 CC FT 551 610
 CC FT 741 779
 CC FT 3 X 13 AA TANDEM REPEATS.
 CC POLY-ARG.

SQ SEQUENCE 992 AA; 109129 MW; 39BEAB9BC515BD84 CRC64;
 Query Match 13.6%; Score 160.5; DB 1; Length 992;
 Best Local Similarity 26.1%; Pred. No. 0.018; Mismatches 62; Conservative 22; Indels 71; Gaps 8;

QY 33 PLHPS-LNIPYGINLPPPLYRPNVTPSYGN-----TYTDT 70
 Db 464 PAHPTPEMPTVILHQPPV-PKVPVVKTPPPSRRRGACVYDDVIEVIVETSDS 522
 QY 71 GLPSTPWLTSRPGPPVYVHIGFPLATQLVNPLPRGFPVPPSRFSAAPAPPIA 130
 Db 523 S-----SVSQPNKFKRHQDGFQSGRQKRAAPVTSF-----SDTGPPAVGPPA 568
 QY 131 AEPAAAPLTPVVAAPBAPGAPVAAEPAPVGAEPAL-----APVAAEPAAEP-- 183
 Db 569 AGPPAPGPPAPGPPAPGPPAPGPPAPGPPAPGPPAPGPPAPGPPAPGPPAPGPP 628
 QY 184 -----VGVEBPA-----EEPSAPAPATKAPAPAPSPSL 214
 Db 629 VRMFMRERQLPOSTGRKPCQCFWEMRAGRBITOMQEPSSHLQSAQTPTTPPSMAPSV 686

RESULT 14
 ENAM_MOUSE STANDARD; PRT; 1274 AA.
 AC 05196;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enamelin precursor.
 GN ENAM.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Swiss Webster; TISSUE=Enamel epithelium;
 RX MEDLINE=20514993; PubMed=11062988;
 RA Hu C.-C., Simmer J.P., Bartlett J.D., Nanci A., Qian Q., Zhang C.,
 RA Hu O.H., Xue J., Fukae M., Uchida T., McDougall M.;
 RT "Murine enamel:in: CDNA and derived protein sequences.";
 CC Connect. Tissue Res. 39:47-61(1998).
 CC -1- FUNCTION: Involved in the mineralization and structural
 organization of enamel. Involved in the extension of enamel during
 the secretory stage of dental enamel formation.
 CC -1- SUBCELLULAR LOCATION: Secreted. Extracellular matrix.
 CC -1- TISSUE SPECIFICITY: Expressed in developing teeth.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U82698; AAB94312.1; -;
 DR PIR; T37193; T37193.
 DR MGD; MGI:1333772; Enam.
 DR GO; GO:0005578; C:extracellular matrix; ISS.
 DR GO; GO:0030345; F:structural constituent of tooth enamel; ISS.
 DR GO; GO:0002822; P:bone mineralization; ISS.
 DR GO; GO:0042476; P:odontogenesis; ISS.
 KM Biomaterialization; Extracellular matrix; Glycoprotein; Signal.
 FT SIGNAL 39
 FT CHAIN 1
 FT CARBOHYD 130
 FT CARBOHYD 252
 FT CARBOHYD 259
 FT CARBOHYD 269
 FT CARBOHYD 300

FT CARBOHYD 1066
 SQ SEQUENCE 1274 AA; 140954 MW; F9DBD1CC9D327143 CRC64;
 Query Match 13.4%; Score 158; DB 1; Length 1274;
 Best Local Similarity 28.0%; Pred. No. 0.03; Mismatches 75; Conservative 18; Indels 95; Gaps 16;

QY 17 KRPPPIGSDNDNDGHLHPSLN-----IPYG-----I 44
 Db 230 KKHDPKPEDPPPPDPPPSASTNSSTVDANAATGSIPEGANDTSPIGNTGPPNAGNPTV 289
 QY 45 RN--LPPLYY-----RPVNTVPSYG--NTYDTGLPSYP-----WILT-----SPG 83
 Db 290 QNGVFPPPKVNVSGQGVKSIQIPRPSQPNLYENYPVNPYSBRQWQTGTGPRONGPG 349
 QY 84 PPTVTHIRG-----FPLATQLVNPLPRGFPV--PSSRFSAAPAPAP 128
 Db 350 YRNQVERGQWNSFANEKQATRGPNPTYGKPPSPSGVYAGNPHVFGNLLPGPNKP 409
 QY 129 IAAEPAAAPLTP-ATP-----VAAEPAPAGAP-VAAEPAAEP-VAAEP 178
 Db 410 VGANPASNKPFVGANPASNKPFVGANPASNKPFVGANPASNKPFVGANPASNKPF 469
 QY 179 AAAP-VGVEBPAEPPS-PAEPATAPKA 204
 Db 470 ASNKPFIGANPANKPISIGTPANKPS 497

RESULT 15
 CCPA_ACEXY STANDARD; PRT; 353 AA.
 ID CCPA_ACEXY
 AC P37697;
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cellulose complementing protein.
 GN CCPAX.
 OS Acetobacter xylinus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Acetobacteraceae; Gluconacetobacter.
 NCBI_TaxID=28448;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 53582;
 RX MEDLINE=94113945; PubMed=8300521;
 RA Stendel R., Iversen T.-G., Coucheron D.H., Fjerslev B., Blatny J.M.,
 RA Valla S.;
 RT "A new gene required for cellulose production and a gene encoding
 cellulolytic activity in Acetobacter xylinum are colocalized with the
 bcs operon".
 RT J. Bacteriol. 176:665-672(1994).
 RL
 RM [2]
 RM SEQUENCE OF 232-353 FROM N.A.
 RP STRAIN=ATCC 53582;
 RC MEDLINE=91322509; PubMed=1830823;
 RX Saxena I.M., Lin P.C., Brown R.M., Jr.;
 RT "Identification of a new gene in an operon for cellulose biosynthesis
 in Acetobacter xylinum".
 RT Plant Mol. Biol. 16:947-954(1991).
 RL
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M96060; AAA16970.1; -;
 DR EMBL; X54676; -; NOT_ANNOTATED_CDS.
 DR PIR; B36963; B36963.
 DR PIR; B36963; B36963.
 FT VARIANT 232
 FT VARIANT 244 250 R -> Q (IN STRAIN ATCC 53582).
 FT VARIANT 244 250 ARLSPRE -> HASAAG (IN STRAIN ATCC 53582).

SQ **SEQUENCE** **353 AA;** **37396 MW;** **287456CAE5E7FB6F CRC64;**

Query Match	13.1%	Score 154;	DB 1;	Length 353;
Best local similarity	29.9%	Pred No 0	017.	

Matches 52; Conservative 9; Mismatches 73; Indels 40; Gaps 7;

Matches 52; Conservative 9; Mismatches 73; Indels 40; Gaps 7;

```
QY      53 YRPVNTVPSPSGNITYTDTGLBPSYPMLTSFGPPYYTHIRGFPLATQLNVPELPKGGPFFV 112
          ||| : :: ||| : | | |
Db      37 YRP-----FVDRSPDVTGVBE--AVERHDAQAEH---DTAVEGQVTPAP----- 75
```

```
QY      53 YRPVNTVPSPSGNITYTDTGLBPSYPMLTSFGPPYYTHIRGFPLATQLNVPELPKGGPFFV 112
          ||| : :: ||| : | | |
Db     37 YRP-----FVDRSPDVTGVBE--AVERHDAQAEH---DTAVEGQVTPAP----- 75
```

Qy 113 PPSRFFSAAAPAPPIAEPAAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEA 172

Db 76 -----QIAVAPRPVPDPRAIVTETAPRPV--VVSAPVTYEPRAAIVPAEPPVQEA 127

QY 173 PVAEPAAEPVGVTEPAAEPPSPAE-----ATAKPAEPHPSPSLEQA 217

Db 128 PVQAPVPPAP--VPPIAQAPPAAPDPASVPPYANVAAPVPEDEAPVTPPAQA 179

Search completed: January 15, 2004, 15:12:28
Job time : 19 secs

Job time : 19 secs

A/Molecule type: DNA
 A/Residues: 1-395 <WHI>
 A/Cross-references: GB:AE001946; GB:AE000513; NID:G6458655; PIDN:AAFI0516.1; PID:G645866
 A/Experimental source: strain R1
 C/Genetics:
 A/Gene: DR0938
 A/Map position: 1

Query Match 15.8%; Score 186.5; DB 2; Length 395;
 Best Local Similarity 35.0%; Pred. No. 0.0001;
 Matches 57; Conservative 11; Mismatches 64; Indels 31; Gaps 9;

QY 60 PSYV-GNTYTDYGLPSYFWILTSFGPYVYHIRGFPLATQLVNPLPRGFPVPSRFF 118
 DB 53 PAEEVGAPOPTTSIPVSGAVADGSPQT---AAPAOQTGAKIPPA--PAAP----- 102
 QY 119 SAAAPAAP-----IAEPAAAPLTATPVAAPAGAPVPAAPAAAPVGAAPAA 170
 DB 103 ---APKIPPEVLDEPRMRPPTPPRPVEETTTAASPTQPPVTAQATQTPPTQTPAA 158
 QY 171 EAPVAAEPAAAPVGVPEPAA-EESPAA---EPAT-AKPAAP 208
 DB 159 QAPATQPPATPAP---EPAAEPAPATTPPATTPTEPAPAP 198

RESULT 7

E86255

hypothetical protein [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C/Accession: E86255

R/Theologian: A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, V.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luo, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: E86255

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-744 <STO>

A/Cross-references: GB:AE005172; NID:G3157926; PIDN:AACT17609.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

Query Match 15.5%; Score 183; DB 2; Length 744;
 Best Local Similarity 28.9%; Pred. No. 0.00031;
 Matches 55; Conservative 14; Mismatches 93; Indels 28; Gaps 6;

QY 33 PLHPSLNIPIGIRNLPPPLYRPVNTVPSYFGNTYDTGLPSY-----FWILTSBGP 85
 DB 461 PPSPPSPPPYVSSPPPYVY---SSPPPPYVSSPPPPYVSSPPPPYVSSPPPPY 516
 QY 86 YVYHIRFPLATQLVNPLPRGFPVPSRFSAAAAPAA--PPIAAAPAAAPLTATP 143
 DB 517 YVY-----SSPPPP--PSPPPPCPSPPPPVYVYVQSPSPSPSPYVYPP 562
 QY 144 VAAEPAGAAVAAEPAAAPVGAEPAAAPVAAEPAAAPVGAEPAAEPSPAPAAAP 203
 DB 563 VTQSPSPSPSPYVYVQSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 622
 QY 204 AAEPAPSPS 213
 DB 623 VTSPSP-PPPS 631

RESULT 8

A40215

Tcd antigen - Trypanosoma cruzi

C/Species: Trypanosoma cruzi

C/Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999

C/Accession: A40215

R/Burns Jr., J.M.; Shreffler, W.G.; Rosman, D.E.; Sleeth, P.R.; March, C.J.; Reed, S.G.

Proc. Natl. Acad. Sci. U.S.A. 89, 1239-1243, 1992

A/Title: Identification and synthesis of a major conserved antigenic epitope of Trypano

A/Reference number: A40215; MUID:92159024; PMID:1371355

A/Accession: A40215

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-207 <BUR>

A/Note: sequence extracted from NCBI backbone (NCBIN:82113, NCBI:P:82117)

C/Superfamily: histone H1

Query Match 15.2%; Score 179; DB 2; Length 207;
 Best Local Similarity 41.0%; Pred. No. 0.00016;
 Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSRFFSMAAPAPPIAEPAAAPLTATPVAAPAGAPVAAEPAAEPVGAEPAAAP 173
 DB 5 PAEPKSAEPKAPK-SABPKAPBSABPKAPBSAGPKAPBSABPKAPBSABPKAP 63
 QY 174 VAAEPAAAPVGVPEPAAEPSP-----AEPATKAPAAEPSPS 213
 DB 64 KPAPKSAEPKAPBSABPKAPBSABPKAPBSABPKAPBSABPKAPBS 108

RESULT 9

T43226

translation initiation factor IF-2 - Stigmatella aurantiaca

C/Species: Stigmatella aurantiaca

C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 02-Feb-2001

C/Accession: T43226

R/Bremund, L.; Ismail, S.; Derjard, B.; Cenciempo, Y.

U. Bacteriol. 179, 2348-2355, 1997

A/Title: Translation initiation factor IF2 of the myxobacterium Stigmatella aurantiaca:

A/Reference number: Z22352; MUID:97234648; PMID:9079922

A/Accession: T43226

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1054

A/Cross-references: EMBL:X87940; PIDN:CAA61162.1

A/Experimental source: strain DW4

C/Genetics:

A/Gene: infB

C/Superfamily: translation initiation factor IF-2; translation elongation factor Tu hom

C/Keywords: nucleotide binding; P-loop; protein biosynthesis

F/555-664/Domain: translation elongation factor Tu homology <ETU>

F/561-568/Region: nucleotide-binding motif A (P-loop)

Query Match 14.9%; Score 175; DB 2; Length 1054;
 Best Local Similarity 38.3%; Pred. No. 0.0013;
 Matches 49; Conservative 8; Mismatches 51; Indels 20; Gaps 4;

QY 96 ATQANVPPLPRGFPVPSRFSAAAAPAPIAA--EPAAAPLTAT--PVAAAPAG 151
 DB 98 ASDVSSPPSP-----VHBSAGAAASBSRVPAALVQGEVAAAPAAASBPAAAPKAT 152
 QY 152 APVAAEPAAAPVGAEPAAAPVAAEPAAAPVGAEPAAAPVGAEPAAEPSP 200
 DB 153 APVAAEPVTEAPKAAAPVAAEPVTEAPKAAAPVAAEPVTEAPKAAAPVTEAP 212
 QY 201 AKPAAP 208
 DB 213 CKGAAPLP 220

RESULT 10

Q03B8

BpLF1 protein - human herpesvirus 4 (strain B95-8)

C/Species: human herpesvirus 4, Epstein-Barr virus

C/Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 16-Jul-1999


```

Db      16 GVDDDTAVVMDSPPI-ASGPPPIEVSAGAPVXSEHAKRPIERQAASTGNSPISTTAEEA 74
Qy      71 GLPSYPMILTSFGFPYVYHIRGFPLATQ-----LNVPLPRPGFPFVPP 114
       75 KVPSSLERGSGSPQPEFTWIKPAPVABSGTDSKADPDASTAAVLQIGPEEVGVFTMTPL 134
Qy      115 SREFFSAAAAPAPPIAEP-AAAAPLTATPVAA-----EPAAAPVAAEPAAEPVGAEP 169
Db      135 D--LPPASDAGDVAABPDGCTAP--ATPASEDNREPPAAA--AAEPAAEP--AAEP 186
Qy      170 AE--APVAAEPAAEA-----FVGEPPAAEPSPAPATAPKAPAEPPSP 212
Db      187 AEPAAEPAAEPAAEAAPVDTFASASGAVPDQEPAPAAAASATPAPAPAAAP 239

```

RESULT 14

```

arylesterase-related protein [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: G87675
R.Nierman, W.C.; Feildbyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gilm, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; NCID:21173698; PMID:11259647
A:Accession: G87675
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-438 <STO>
A:Cross-references: GB:AE005673; NID:g13425157; PIDN:AAK25403.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC3441

```

```

Query Match      14.6%; Score 171.5; DB 2; Length 438;
Best Local Similarity 33.3%; Pred. No. 0.00094;
Matches 68; Conservative 14; Mismatches 59; Indels 63; Gaps 11;

Qy      64 GNTYTDGTPSYR---WLTSGRF-----PYVYHIRGFPLATQNLVPLPRPGFPFVPP 114
Db      219 GKRTASADIKIYPNNNAHYLIDEPNTMKIIDDVIAWLRGKTQAPAACTVBPAP----- 270
Qy      115 SREFFSAAAAP-APPIAEPAAAPLTATPV--AAEPAA-----GAPVAAEP-----A 160
Db      271 ----PAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 323
Qy      161 EAPVG-----AEPAAEPV--AAEPAAEPVGEPPAAEPSPAP----- 197
Db      324 KAPVAKKAPAKKAAAPAAAPAKSAAAPKAKAPAAKAAAPKAAAPKAAKAPKAAKAPAAE 363
Qy      198 --PATKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
Db      384 TAPAAKAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPATK 407

```

RESULT 15

```

T17547
Proline-rich protein A57R - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17547
R.Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17547
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAK96425.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:

```

```

A:Gene: A57R
Query Match      14.5%; Score 171; DB 2; Length 544;
Best Local Similarity 27.8%; Pred. No. 0.0012;
Matches 58; Conservative 19; Mismatches 106; Indels 26; Gaps 4;

```

```

Qy      2 KILLWACIVCAFAKRRP---PTGEDNDGDHPLHPLINIPYGRNLPPPLYRPNVT 58
Db      4 RVIFPLVLTAFAPABEPTTQIVTQTTVNSPKPSPKPSKPSKPSKPSKPSKPSKPSKPSKPS 63
Qy      59 VPSYQNTYTDGLPSYPMILTSFGFPYVYHIRGFPLATQNLVPLPRPGFPFVPPSPRP 118
Db      64 KPPSPKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPS 100
Qy      119 SAAAAPAPPIAEPAAAPLTATPVAAEPAAAPVAAEPAAAPVGAEPAAEPVAAEP 178
Db      101 PSPKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPSKPS 160
Qy      179 AAAPVGEPPAAEPSPAPATAPKAPAP 207
Db      161 PAPBPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 189

```

Search completed: January 15, 2004, 15:14:44
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004. CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 15:14:15 ; Search time 33 Seconds
(without alignments)
1357.012 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178

Sequence: 1 MKLLWACIVCAVFAKRRF.....TAKPAEPHPSLEQANQ 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 20481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1178	100.0	219	9 US-09-922-480-2	Sequence 2, Appl1
2	1178	100.0	219	9 US-09-923-236-2	Sequence 2, Appl1
3	1178	100.0	219	10 US-09-922-469-2	Sequence 2, Appl1
4	1172	99.5	219	11 US-09-746-783-106	Sequence 106, App
5	1113.5	94.5	221	12 US-10-231-417-192	Sequence 192, App
6	181	15.4	2657	12 US-10-316-253-95	Sequence 95, Appl
7	175.5	14.9	903	12 US-10-264-049-2409	Sequence 2409, Ap
8	173	14.7	538	14 US-09-976-740-43	Sequence 43, Appl
9	173	14.7	538	14 US-10-023-523-43	Sequence 43, Appl
10	173	14.7	538	14 US-10-023-523-43	Sequence 43, Appl
11	173	14.7	538	14 US-10-029-386-33686	Sequence 33686, A
12	171.5	14.6	693	12 US-10-108-2604-3849	Sequence 3849, Ap
13	170	14.4	267	12 US-10-084-843-142	Sequence 142, App
14	170	14.4	267	12 US-10-193-002-137	Sequence 137, App
15	170	14.4	428	10 US-09-906-514-4	Sequence 4, Appl1

16	170	14.4	428	12 US-10-177-478-14	Sequence 14, Appl
17	169.5	14.4	802	9 US-09-823-240-2	Sequence 2, Appl1
18	168.5	14.3	309	11 US-09-820-843A-24	Sequence 24, Appl
19	168.5	14.3	550	10 US-09-976-740-47	Sequence 47, Appl
20	168.5	14.3	550	14 US-10-023-523-47	Sequence 47, Appl
21	168.5	14.3	550	14 US-10-023-523-47	Sequence 47, Appl
22	166	14.1	419	15 US-10-156-761-13549	Sequence 13549, A
23	162	13.8	98	11 US-09-820-843A-85	Sequence 85, Appl
24	159.5	13.5	527	9 US-09-888-615-104	Sequence 104, App
25	159	13.5	1276	15 US-10-156-761-10509	Sequence 10509, A
26	157	13.3	1479	15 US-10-156-761-11739	Sequence 11739, A
27	156.5	13.3	800	12 US-10-029-386-332198	Sequence 32198, A
28	156	13.2	3122	12 US-10-200-562-201	Sequence 201, App
29	156	13.2	3122	12 US-10-237-551-201	Sequence 201, App
30	156	13.2	3122	12 US-10-237-551-250	Sequence 250, App
31	156	13.2	4952	12 US-10-051-874-56	Sequence 56, Appl
32	156	13.2	5008	12 US-10-051-874-166	Sequence 166, App
33	155	13.2	574	12 US-10-168-097A-76	Sequence 76, Appl
34	155	13.2	574	12 US-10-239-431A-38	Sequence 38, Appl
35	155	13.2	639	12 US-10-369-493-2275	Sequence 2275, Ap
36	154.5	13.1	433	10 US-09-906-514-2	Sequence 2, Appl1
37	152.5	12.9	1387	15 US-10-156-761-13000	Sequence 13000, A
38	151	12.8	639	12 US-10-369-493-3962	Sequence 3962, App
39	150	12.7	352	11 US-09-820-843A-23	Sequence 23, Appl
40	149	12.6	981	12 US-10-029-386-33035	Sequence 33035, A
41	148.5	12.6	259	12 US-10-029-386-31968	Sequence 31968, A
42	148	12.6	582	15 US-10-156-761-11938	Sequence 11938, A
43	148	12.6	706	12 US-10-369-493-11953	Sequence 11953, A
44	148	12.6	2971	15 US-10-146-473-50	Sequence 50, Appl
45	147	12.5	199	15 US-10-034-934-125	Sequence 125, App

ALIGNMENTS

RESULT 1
US-09-922-480-2
; Sequence 2, Application US/09922480
; Patent No. US20020081701A1
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/922,480
; CURRENT FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-922-480-2

Query Match	100.0%	Score 1178;	DB 9;	Length 219;
Best Local Similarity	100.0%;	Pred. No. 2.6e-77;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKLLWACIVCAVFAKRRFPFGIEDNDGHLHPSLNIPIYGRILAPPLYRPNVTVP	60	
Db	1	MKLLWACIVCAVFAKRRFPFGIEDNDGHLHPSLNIPIYGRILAPPLYRPNVTVP	60	
Qy	61	SYPGNTYDTGLPSYWIITSGFPYVYHIRGFLATQINVPPLPRGPFVPSRFFSA	120	
Db	61	SYPGNTYDTGLPSYWIITSGFPYVYHIRGFLATQINVPPLPRGPFVPSRFFSA	120	
Qy	121	AAAPAPAPTAAPPAAPLTATPVAEPAAGAPVAEPAABAPVGAEPAAEPAAPAA	180	
Db	121	AAAPAPAPTAAPPAAPLTATPVAEPAAGAPVAEPAABAPVGAEPAAEPAAPAA	180	
Qy	181	EAPVGEPAABEPSPAEPAATAPPAEPPHPSLEQANQ	219	

Db 181 EAPVGEPAAEPSPAEPATAKPAABEPHPSLSLEQANQ 219

RESULT 2
US-09-923-236-2
Sequence 2, Application US/09923236

Patent No. US20020090677A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/923,236
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-923-236-2

Query Match 100.0%; Score 1178; DB 9; Length 219;
Best Local Similarity 100.0%; Pred. No. 2,6e-77;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

DB 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

QY 61 SYPGNTYDTGTPSYWMLTSPGFPYVYHNGFPLATOLNVPPLPRGFPVPSRFFSA 120

DB 61 SYPGNTYDTGTPSYWMLTSPGFPYVYHNGFPLATOLNVPPLPRGFPVPSRFFSA 120

QY 121 AAAPAPPIAABPAAPLTPVAABPAAGAVAAEPAAEAPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAPPIAABPAAPLTPVAABPAAGAVAAEPAAEAPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAABEPHPSLSLEQANQ 219

DB 181 EAPVGEPAAEPSPAEPATAKPAABEPHPSLSLEQANQ 219

RESULT 3
US-09-922-469-2

Sequence 2, Application US/09922469
Patent No. US20020173027A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OF INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,469
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-469-2

Query Match 100.0%; Score 1178; DB 10; Length 219;
Best Local Similarity 100.0%; Pred. No. 2,6e-77;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

Db 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

QY 61 SYPGNTYDTGTPSYWMLTSPGFPYVYHNGFPLATOLNVPPLPRGFPVPSRFFSA 120

DB 61 SYPGNTYDTGTPSYWMLTSPGFPYVYHNGFPLATOLNVPPLPRGFPVPSRFFSA 120

QY 121 AAAPAPPIAABPAAPLTPVAABPAAGAVAAEPAAEAPVGAEPAAEPVAAEPAA 180

DB 121 AAAPAPPIAABPAAPLTPVAABPAAGAVAAEPAAEAPVGAEPAAEPVAAEPAA 180

QY 181 EAPVGEPAAEPSPAEPATAKPAABEPHPSLSLEQANQ 219

DB 181 EAPVGEPAAEPSPAEPATAKPAABEPHPSLSLEQANQ 219

RESULT 4
US-09-746-783-106
Sequence 106, Application US/09746783

Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth

McCoy, John M.
Lavallee, Edward R.
Racie, Lisa A.
Treacy, Maurice
Spaulding, Vikki
Agostino, Michael J.
Howes, Steven H.
Fechtel, Kim

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM

NUMBER OF SEQUENCES: 231

CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA

COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/746,783

FILING DATE: 21-Dec-2000

CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Mlaelincic, Debra J.

REGISTRATION NUMBER: 46,931

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids

TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 106:

US-09-746-783-106

Query Match 99.5%; Score 1172; DB 11; Length 219;
Best Local Similarity 99.1%; Pred. No. 7e-77;
Matches 217; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

DB 1 MLLLMACIVCAFAAKRRPPFPGEDNDGHPHLSINIPYGINLPPLPYRRPVNTVP 60

QY 61 SYPGNTYDTGTPSYWMLTSPGFPYVYHNGFPLATOLNVPPLPRGFPVPSRFFSA 120

```

Db      61 SYGNTYTDGLSPWILTSPPGPPVYHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Qy      121 AAAPAPPIAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 180
Db      121 AAAPAPPIAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 180
Qy      181 EAPVGEPAEAPSPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
Db      181 EAPVGEPAEAPSPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219

```

RESULT 5

```

US-10-231-417-192
; Sequence 192, Application US/10231417
; Publication No. US20030176681A1
; GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019P1
; CURRENT APPLICATION NUMBER: US/10/231,417
; PRIOR FILING DATE: 2002-08-30
; PRIOR APPLICATION NUMBER: US/09/296,622
; NUMBER OF SEQ ID NOS: 619
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 192
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: SITE
; LOCATION: (159)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (221)
; OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-192

```

Query Match 94.5%; Score 1113.5; DB 12; Length 221;

Best Local Similarity 95.0%; Pred. No. 1.1e-72; Matches 209; Conservative 1; Mismatches 9; Indels 1; Gaps 1;

```

Qy      1 MKLLLAACIVCAFAKRRRPPPIGDDNDGHPHLSINIPYGRMLPPPLYRPNVTV 60
Db      1 MKLLLAACIVCAFAKRRRPPPIGDDNDGHPHLSINIPYGRMLPPPLYRPNVTV 60
Qy      61 SYGNTYTDGLSPWILTSPPGPPVYHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Db      61 SYGNTYTDGLSPWILTSPPGPPVYHIRGFPPLATQLVNPPRPGPFVPSRFFSA 120
Qy      121 AAAPAPPIAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 179
Db      121 AAAPAPPIAEPAAAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAA 180
Qy      180 AEAAPVGEPAEAPSPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 219
Db      181 AEAAPVGEPAEAPSPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 220

```

RESULT 6

```

US-10-316-253-95
; Sequence 95, Application US/10316253
; Publication No. US20030162706A1
; GENERAL INFORMATION:
; APPLICANT: The Procter & Gamble Company
; APPLICANT: Peters, Kevin
; APPLICANT: Thompson, Larry
; APPLICANT: Wang, Feng
; APPLICANT: Greis, Kenneth
; TITLE OF INVENTION: Angiogenesis Modulating Proteins
; FILE REFERENCE: 8865M

```

```

; CURRENT APPLICATION NUMBER: US/10/316,253
; PRIOR FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: US 60/355,295
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 308
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 95
; LENGTH: 2657
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-316-253-95

```

Query Match 15.4%; Score 181; DB 12; Length 2657;

Best Local Similarity 29.1%; Pred. No. 0.00012; Matches 57; Conservative 22; Mismatches 61; Indels 56; Gaps 8;

```

Qy      40 IYGRIRNLPP-----LYRPVNTVPSFGNTYTDGLSPWILTSPPGPPVHIRGF 93
Db      2239 VRFGRKQINPPTANSLSKVVTTIK--PSNSSKPTAL-----VNLTP----- 2280
Qy      94 PLATQLVNPPRPGPFVPSRFFSAAPAPAPAPAPAPAPAPAPAPAPAPAPAP 133
Db      2281 -----AKPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 2333
Qy      134 AAAPAPLTATVAAEPAAAGAVAAEPAAEPAAEPAAEPAAEPAAEPAAEPAA 192
Db      2334 ASVRPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 2389
Qy      193 PSPAPATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 208
Db      2390 PMPAPVLTSAAPVR 2405

```

RESULT 7

```

US-10-264-049-2409
; Sequence 2409, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2409
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: MISC_FEATURE
; LOCATION: (10)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (15)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (16)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
; NAME/KEY: MISC_FEATURE
; LOCATION: (795)
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2409

```

Query Match 14.9%; Score 175.5; DB 12; Length 903;

Best Local Similarity 32.3%; Pred. No. 9.9e-05;

TYPE: PRT
ORGANISM: Homo sapiens
US-10-023-523-43

Query Match 14.7%; Score 173; DB 14; Length 538;
Best Local Similarity 36.8%; Pred. No. 8.9e-05;
Matches 49; Conservative 7; Mismatches 47; Indels 30; Gaps 5;

QY 102 PPLPFRGFVPPSRFPFAAAPPPIAAEPAAALPLTTPVAABEPAGAPVAABEPAAE 161
DB 111 PPAAPRARP-----AAAAAAPPPTPAPPPPPPAVAAAAAPARARAAAAATAPSP 162
QY 162 APVGAEPAAE-APVAA-EPAAEPVGEV-----AAEP-----SPAEP 169
DB 163 GFAOPGFRAGRAAPLAPPPAPAPVAVBPAGFRAPPPAVAAABEPPLPPPPQPPAPQ 222
QY 200 TAKPAABEPHSP 212
DB 223 QQQPPPPQPPQ 235

RESULT 11

US-10-029-386-33686
Sequence 33686, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: AEOMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1
SEQ ID NO 33686
LENGTH: 980
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AL137853.3
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.84
OTHER INFORMATION: SWISSPROT HIT: Q02910, EVALU 2.00e-03
US-10-029-386-33686

Query Match 14.7%; Score 173; DB 12; Length 980;
Best Local Similarity 31.1%; Pred. No. 0.00016;
Matches 68; Conservative 13; Mismatches 112; Indels 26; Gaps 9;

QY 6 MACIVCAVAPARKRRPFIQEDNDGRLHPSLNI-----PYGIRNLPPPLYRRPVT 58
DB 531 WAAALATV-----PTEEDGTREG-PVTPATVHAPEBDTAAVSVTPBEBSAAA 582
QY 59 VPSYGTNTDTGLPSYPMILTSFGPPYVYHNGFPLATQLVNPLPRGPPVP-PSRF 117
DB 583 VPTPEEPTSPAAPVPT-PEEPTSPAAPVPTPEEPTSPAAPVPTPEEPTSPAAPVPTPEE 641
QY 118 FSAAPAAAPPIAAEPAAAPLPTTPVAABEPAGAPVAAE---PAAAPVGAEPAAAPV 174
DB 642 TSPAAAVPTPEEPTSPAAPVPTPEEPTSPAAPVPTPEEPTSPAAPVPTPEEPA--SPA 697
QY 175 AAEPAAAPVGAEPAAEPSPAPAPATKAPABEPHSPS 213
DB 698 AAVPTPEEP--ASPAAVPTPEEPAFAPAPVPTPEEBSAS 734

RESULT 12

US-10-108-260A-3849
Sequence 3849, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: NO. US20040005560A1e1 full length cDNA
FILE REFERENCE: H1-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3849
LENGTH: 693
TYPE: PRT
ORGANISM: Homo sapiens
US-10-108-260A-3849

Query Match 14.6%; Score 171.5; DB 12; Length 693;
Best Local Similarity 31.6%; Pred. No. 0.00015;
Matches 62; Conservative 17; Mismatches 78; Indels 39; Gaps 11;

QY 33 PLHPSLNIPIYGRNLPPPLYRRPVTVPSPGNTYDTGLPSYPMILTSFGPPYVYHNG 92
DB 253 PVSPTSATPP--SQAPPSLAAPPLQVPSPPASP-----PMSPTSATPPQAPP----- 297
QY 93 FLPLAT-QLNVPLPLPPRFPVPPSPRFSAAAP-----AAPPIAAEPAAALPTATPA 145
DB 298 -PLAEPPLQVPPSPSPSPMS-----SATPPRVPLLAAPLQVPSPPASLPSPLA 351
QY 146 AEPAGAPVAAEPAAAPVGAEPAAE---EAPVAAEPAAEPVGEVPAEPSPAPAPTAK 202
DB 352 KPPPQAPPALATPLQA-LPSPPSPGQAPFS--PSASLP--MSPLATPPQAPVTLAA 406
QY 203 P---AAEPHPSPL 215
DB 407 PLLQVPPSPSPSPTLQ 422

RESULT 13

US-10-084-843-142
Sequence 142, Application US/10084843
Publication No. US20030134243A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESSES:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/084,843
FILING DATE: 25-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9

```

1 TELECOMMUNICATION INFORMATION:
2 TELEPHONE: (206) 622-4900
3 TELEFAX: (206) 682-6031
4 INFORMATION FOR SEQ ID NO: 142:
5 SEQUENCE CHARACTERISTICS:
6 LENGTH: 267 amino acids
7 TYPE: amino acid
8 STRANDEDNESS: single
9 TOPOLOGY: linear
10 MOLECULE TYPE: peptide
11 SEQUENCE DESCRIPTION: SEQ ID NO: 142:
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
10
```

```

: REFERENCE/DOCKET NUMBER: 210121.417C9
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (206) 622-4900
: TELEFAX: (206) 682-6031
: INFORMATION FOR SEQ ID NO: 137:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 267 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: SEQUENCE DESCRIPTION: SEQ ID NO: 137:
US-10-193-002-137

Query Match      14.4%; Score 170; DB 12; Length 267;
Best Local Similarity 28.0%; Pred. No. 7,28-05;
Matches 56; Conservative 14; Mismatches 94; Indels 36; Gaps 11;

Cy 21 PFIGEDNDGDHPLPSLNIPIY-GIRNLPRLYYRPVN-TVPSYPGNTYTDGLPSYEW 77
Db 79 PRFSTSGVGSTGPSAPAEAPAVGVPAVPPI--EVPILIPFPFG--WQPMPTIP- 131
Cy 78 ILTSPGPPYYHNRGFLAQQLANPLPRGFPVPPSRFSSAALAPAPIAEPAADA 137
Db 132 -TAP--PTT-----PVTTSATTPPTTPTTPTT-----PTTP--TTVPTTP 170
Cy 138 PLT--ATPVAEPAAGAPVAAPAEAPVGAEEPAEAPVAEPAEAPVGEPAEAPS- 194
Db 171 PTPPTPTPTPTPTPTVAPTTVAPTTVAPTTVAPTTVAPTTVAPTTVAPTTVA 230
Cy 195 --PAEPATKPAEPHPSP 212
Db 231 QMPTQQQTVAAPQTVAAPAP 250

RESULT 15
US-09-906-514-4
: Sequence 4, Application US/09906514
: Patent No. US20020170085A1
: GENERAL INFORMATION:
: APPLICANT: Kaeppler, Shawn
: APPLICANT: Springer, Nathan
: APPLICANT: Phillips, Ronald
: TITLE OR INVENTION: Methyl CpG Binding Domain Nucleic Acids from Maize
: FILE REFERENCE: Methyldinding
: CURRENT APPLICATION NUMBER: US/09/906,514
: NUMBER OF SEQ ID NOS: 13
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 428
: TYPE: PRT
: ORGANISM: Zea mays
US-09-906-514-4

Query Match      14.4%; Score 170; DB 10; Length 428;
Best Local Similarity 33.3%; Pred. No. 0,00012;
Matches 74; Conservative 13; Mismatches 89; Indels 46; Gaps 11;

Cy 25 EDDNDGDHPLPSLNIPIYGIRNLPRLYYRPVNTVPSYPGNTYTDGLPSYEWILTSBGF 84
Db 183 EEKEDAKRDEPEAVADAPVGN--PTENSAPAPAEPAVPAVPETBSVAEPAVLAPA- 239
Cy 85 PYYHNGFPLAQQLANPLPRGFPVPPSRFSSAAL-----PAAPPI----- 129
Db 240 PETH-----PAKPAAPAPAPAPENKEDAP-----AAAAAPVDTKSAEPPAAAPDTKS 290
Cy 130 AAEPAALAAFLTATPVAEPAAGAPVAAPAEAPVGAEEPAEPAEAV-----AAEPAEAPV 184
Db 291 VAEEPAALAAFLVETTLVAESAADA--VAAPAPETKSDAEPAAP-AVVPETKPVAAESAADA 347
Cy 185 GVEPAEAPSPAPAPATKPAAP-----PFPSSGLEDA 217

```

Tue Jan 20 17:43:12 2004

us-09-923-236-2.rapb

Page 7

Db 348 A--PAPETKSDAEPAAAAAPAPETKSDAAAAADPAFGTKADAA 387

Search completed: January 15, 2004, 15:19:47
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:11:59 ; Search time 21 Seconds
(without alignments)
441.242 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1178

Sequence: 1 MKLLMAGICVAFARRRRP.....TAKPAAPRPPSPSLQANQ 219

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1178	100.0	219	US-09-527-345-2	Sequence 2, Appl1
2	1179	15.2	207	US-07-836-642-2	Sequence 2, Appl1
3	1179	15.2	207	US-08-169-563-2	Sequence 2, Appl1
4	1179	15.2	207	US-08-403-379A-6	Sequence 6, Appl1
5	1179	15.2	207	US-08-557-309B-50	Sequence 50, Appl1
6	1179	15.2	207	US-08-929-414-6	Sequence 6, Appl1
7	1179	15.2	207	US-08-834-306-50	Sequence 50, Appl1
8	1179	15.2	207	US-08-993-674A-50	Sequence 50, Appl1
9	1179	15.2	207	US-09-256-976-50	Sequence 50, Appl1
10	1179	14.4	267	US-08-818-112-142	Sequence 142, App
11	1170	14.4	267	US-08-818-111-137	Sequence 137, App
12	1170	14.4	267	US-09-056-556-142	Sequence 142, App
13	1170	14.4	267	US-09-072-596-137	Sequence 137, App
14	164.5	14.0	316	US-09-252-991A-32957	Sequence 32957, A
15	163.5	13.9	214	US-08-217-327-4	Sequence 4, Appl1
16	154	13.1	805	US-09-103-429A-4	Sequence 4, Appl1
17	153	13.0	786	US-09-103-429A-3	Sequence 3, Appl1
18	150	12.7	399	US-09-252-991A-22853	Sequence 22853, A
19	148	12.6	2972	US-09-579-181-2	Sequence 2, Appl1
20	148	12.6	3118	US-09-579-181-1	Sequence 1, Appl1
21	145.5	12.4	1565	US-09-252-991A-20967	Sequence 20967, A
22	144	12.2	195	US-09-252-991A-27942	Sequence 27942, A
23	142.5	12.1	515	US-09-252-991A-25394	Sequence 25394, A
24	141	12.0	369	US-07-757-022B-14	Sequence 14, Appl1
25	139.5	11.8	941	US-07-757-022B-84	Sequence 84, Appl1
26	139.5	11.8	1022	US-07-757-022B-74	Sequence 74, Appl1
27	139.5	11.8	1038	US-07-757-022B-74	Sequence 74, Appl1

28	139.5	11.8	1049	US-07-757-022B-58	Sequence 58, Appl1
29	139.5	11.8	1140	US-07-757-022B-104	Sequence 104, Appl1
30	139.5	11.8	1270	US-07-757-022B-44	Sequence 44, Appl1
31	139.5	11.8	1311	US-07-757-022B-42	Sequence 42, Appl1
32	139.5	11.8	1313	US-07-757-022B-142	Sequence 142, Appl1
33	139.5	11.8	1314	US-07-757-022B-50	Sequence 50, Appl1
34	139.5	11.8	1320	US-07-757-022B-46	Sequence 46, Appl1
35	139.5	11.8	1330	US-07-757-022B-60	Sequence 60, Appl1
36	139.5	11.8	1354	US-07-757-022B-40	Sequence 40, Appl1
37	139.5	11.8	1361	US-07-757-022B-52	Sequence 52, Appl1
38	139.5	11.8	1363	US-07-757-022B-52	Sequence 52, Appl1
39	139.5	11.8	1404	US-07-757-022B-2	Sequence 2, Appl1
40	139.5	11.8	1404	US-07-757-022B-62	Sequence 62, Appl1
41	138	11.7	180	US-07-757-022B-7	Sequence 7, Appl1
42	138	11.7	180	US-07-757-022B-6	Sequence 6, Appl1
43	138	11.7	304	US-09-252-991A-23116	Sequence 23116, A
44	136.5	11.6	504	US-09-219-849-3	Sequence 3, Appl1
45	136.5	11.6	561	US-08-642-255-52	Sequence 52, Appl1

ALIGNMENTS

```

RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Adler, David A.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; CURRENT FILING DATE: 1999-03-17
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match      100.0%; Score 1178; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 3.3e-93;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MKLLMAGICVAFARRRRPFIQEDDNDGHLPSLNIPIGIRMLPPPLYRPNVTVP 60
DB      1 MKLLMAGICVAFARRRRPFIQEDDNDGHLPSLNIPIGIRMLPPPLYRPNVTVP 60

QY      61 SYGNTYTDGLPSYRWILTSQGFPPVYHIRGFLATQNLNVPPLPRGPFVPSRFFSA 120
DB      61 SYGNTYTDGLPSYRWILTSQGFPPVYHIRGFLATQNLNVPPLPRGPFVPSRFFSA 120

QY      121 AAPAPPIAEPAPAAPIATATVAAPAAAGVAAEPAAAEVGAEPAAEAEPAA 180
DB      121 AAPAPPIAEPAPAAPIATATVAAPAAAGVAAEPAAAEVGAEPAAEAEPAA 180

QY      181 EAPVGEPAAPAEPSPAEPATAKPADEPHSPSLQANQ 219
DB      181 EAPVGEPAAPAEPSPAEPATAKPADEPHSPSLQANQ 219

RESULT 2
US-07-836-642-2
; Sequence 2, Application US/07836642
; Patent No. 5304371
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
; AGAINST T. cruzi Infection

```

NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-836-642-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAPAPAPPIAEPAAAPLTATPVAAEPBAGPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGVPEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 3
US-08-169-563-2
Sequence 2, Application US/08169563
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: Peptide for Diagnosing and Immunizing
TITLE OF INVENTION: Against T. cruzi Infection
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Apple Operating System 7.1
SOFTWARE: Microsoft Word for Apple, Version 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/169,563
FILING DATE: 17-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/836,642
FILING DATE: 14-FEB-1992
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia A.
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: REED-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206)587-0430
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-169-563-2

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAPAPAPPIAEPAAAPLTATPVAAEPBAGPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEKPAEPKSAEPKPAEPKSAEPKPAEPKSAEPKPAEPKSAEP 63
QY 174 VAAEPAAEPVGVPEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 4
US-08-403-379A-6
Sequence 6, Application US/08403379A
Patent No. 5756662
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6100 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,379A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-403-379A-6

Query Match 15.2%; Score 179; DB 1; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;
QY 114 PSRFSAAAPAPAPPIAEPAAAPLTATPVAAEPBAGPVAAEPAAEPVGAEPAAEP 173

Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKAGFPKPAEPKSAEPKPAEPKSAEP 63
Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 5

US-08-557-309B-50
Sequence 50, Application US/08557309B
Patent No. 5915572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-557-309B-50

Query Match 15.2%; Score 179; DB 2; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFFSAAAPAPAPPIAEPKAAAPLTATVVAEPKAGAPVAEPKSAEPKSAEPKSAEP 173
Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 6

US-08-929-414-6
Sequence 6, Application US/08929414
Patent No. 5942403
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Haughton, Raymond
APPLICANT: Skeiky, Yasir A.W.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OPT. CRUZI INFECTION
NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/929,414
FILING DATE: 15-SEP-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-929-414-6

Query Match 15.2%; Score 179; DB 2; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

Qy 114 PSRFFSAAAPAPAPPIAEPKAAAPLTATVVAEPKAGAPVAEPKSAEPKSAEPKSAEP 173
Db 5 PAEPKSAEPKPAEPK-SABEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
Qy 174 VAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 213
Db 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 7

US-08-834-306-50
Sequence 50, Application US/08834306
Patent No. 6054135
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 8

US-08-993-674A-50
Sequence 50, Application US/08993674A
Patent No. 6228372
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 81
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 207 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-50

Query Match 15.2%; Score 179; DB 3; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 9

US-09-256-976-50
Sequence 50, Application US/09256976
Patent No. 6419933
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
APPLICANT: Smith, John M.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
FILE REFERENCE: 210121.422C3
CURRENT APPLICATION NUMBER: US/09/256,976
CURRENT FILING DATE: 1999-02-24
NUMBER OF SEQ ID NOS: 95
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 50
LENGTH: 207
TYPE: PRT
ORGANISM: Trypanosoma cruzi
US-09-256-976-50

Query Match 15.2%; Score 179; DB 4; Length 207;
Best Local Similarity 41.0%; Pred. No. 4.1e-08;
Matches 43; Conservative 11; Mismatches 45; Indels 6; Gaps 2;

QY 114 PSHFSAAPAAAPPIAEPAAAPLTATPVAAEPAGAPVAAEPAAEPVGAEPAAEP 173
DB 5 PAEPKSAEPKPAEPK-SAEPPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 63
QY 174 VAAEPAAEPVGEPAEPSP-----AEPATKPAEPHPSPS 213
DB 64 KPAEPKSAEPKPAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEPKSAEP 108

RESULT 10

US-08-818-112-142
Sequence 142, Application US/08818112
Patent No. 6230969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Melo, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twadzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS


```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,327
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/812,233
; FILING DATE: 19-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 112290831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 214 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-217-327-4

Query Match 13.9%; Score 163.5; DB 1; Length 214;
Best Local Similarity 30.3%; Pred. No. 8.9e-07;
Matches 67; Conservative 20; Mismatches 81; Indels 53; Gaps 14;

QY 2 KLLIMACIVCAFAKRRPPPIGEDNDGHPHAPSINIPYGIKRLPPPLYRPPVNTVPS 61
Db 6 KNLEFLSALLCIAVA-----GVLGQ-----APSNPRTSTP-----ATPTP-----PASTPPP 46
QY 62 YPGNTYTDGLPSYPMILTSPPGYVYHNGFPIATQLNVPL---PRGFPPVPPSRP 117
Db 47 TTAAPPPTIATP--PVTSTP-----PTSSP---PVTASPPVSTP--PPS-- 86
QY 118 FSAAPAAAPAPPIAAEPAAAPLVTATPVAAEPAAAGAPVAAEPAAAPVGAEPAAAPVAAE 177
Db 87 ---SPPATPPAPSPPTPPAPSPPTPPAPSPPTPPAPSPPTPPAPSPPTPPAPSPPTPPA-- 141
QY 178 PAEAPVGEPAEAPSPAP--PATAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 217
Db 142 PLASPPATV--PAT---SPVQTPLTSPAPAPPTAPAPPTLGA 178
```

Search completed: January 15, 2004, 15:15:17
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:29:52 ; Search time 33 Seconds

(without alignments)
1357.012 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 1 MKLLWACTIVCAVAFARRRRF.....TAKRAPEHPSPLEQANQ 219

Sequence: 1

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 762491 seqs, 204481190 residues

Word size: 0

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	9 US-09-922-480-2	Sequence 2, Appl1
2	219	100.0	219	9 US-09-923-236-2	Sequence 2, Appl1
3	219	100.0	219	10 US-09-922-469-2	Sequence 192, App
4	150	68.5	221	12 US-10-231-417-192	Sequence 106, Appl
5	140	63.9	219	11 US-09-746-783-106	Sequence 1363, A
6	9	4.1	798	15 US-10-156-761-13162	Sequence 4434, Ap
7	9	4.1	2969	10 US-09-738-626-4434	Sequence 56, Appl
8	8	3.7	132	12 US-10-289-762-96	Sequence 250, App
9	8	3.7	206	15 US-10-219-220-250	Sequence 82, Appl
10	8	3.7	249	15 US-10-180-375-82	Sequence 12383, A
11	8	3.7	342	12 US-10-369-493-12383	Sequence 7317, Ap
12	8	3.7	367	12 US-10-369-493-7317	Sequence 10, Appl
13	8	3.7	383	10 US-09-788-345-10	Sequence 12, Appl
14	8	3.7	383	12 US-10-337-312-10	Sequence 12, Appl
15	8	3.7	412	10 US-09-788-345-12	

16	8	3.7	412	12 US-10-337-312-12	Sequence 12, Appl
17	8	3.7	428	10 US-09-906-514-4	Sequence 4, Appl1
18	8	3.7	428	12 US-10-177-478-14	Sequence 14, Appl1
19	8	3.7	454	15 US-10-156-761-13939	Sequence 13939, A
20	8	3.7	511	9 US-09-864-761-14590	Sequence 34590, A
21	8	3.7	545	12 US-10-929-386-32280	Sequence 32280, A
22	8	3.7	695	12 US-10-369-493-15364	Sequence 15364, A
23	8	3.7	850	12 US-10-369-493-15364	Sequence 10111, A
24	8	3.7	4307	12 US-10-369-493-5698	Sequence 5698, Ap
25	8	3.7	4307	12 US-10-369-493-5698	Sequence 5698, Ap
26	8	3.7	4307	12 US-10-369-493-5700	Sequence 5700, Ap
27	8	3.7	8026	12 US-10-132-134-12	Sequence 12, Appl
28	8	3.2	16	12 US-10-397-551-43	Sequence 43, Appl
29	7	3.2	21	14 US-10-124-800-15	Sequence 15, Appl
30	7	3.2	27	11 US-09-974-879-306	Sequence 306, Appl
31	7	3.2	27	11 US-09-305-736-306	Sequence 306, Appl
32	7	3.2	27	12 US-09-818-683-306	Sequence 306, Appl
33	7	3.2	48	9 US-09-739-907-177	Sequence 177, Appl
34	7	3.2	48	12 US-09-938-671-177	Sequence 177, Appl
35	7	3.2	60	9 US-09-864-761-33554	Sequence 33554, A
36	7	3.2	60	9 US-09-864-761-33624	Sequence 33624, A
37	7	3.2	60	9 US-09-864-761-34127	Sequence 34127, A
38	7	3.2	61	9 US-09-864-761-39187	Sequence 39187, A
39	7	3.2	65	9 US-09-864-761-36194	Sequence 36194, A
40	7	3.2	78	11 US-09-820-843A-89	Sequence 89, Appl
41	7	3.2	88	9 US-09-764-853-503	Sequence 503, Appl
42	7	3.2	91	9 US-09-867-550-738	Sequence 738, Appl
43	7	3.2	95	9 US-09-739-907-178	Sequence 178, Appl
44	7	3.2	95	12 US-09-938-671-178	Sequence 178, Appl
45	7	3.2	105	12 US-10-104-047-2874	Sequence 2874, Ap

ALIGNMENTS

RESULT 1
US-09-922-480-2
Sequence 2, Application US/09922480
Patent No. US20020081701A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Adler, David A.
TITLE OR INVENTION: SECRETED SALIVARY ZS1G63 POLYPEPTIDE
FILE REFERENCE: 97-71
CURRENT APPLICATION NUMBER: US/09/922,480
CURRENT FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 60/124,820
PRIOR FILING DATE: 1999-03-17
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 219
TYPE: PRT
ORGANISM: Homo sapiens
US-09-922-480-2

Query Match	100.0%	Score 219	DB 9	Length 219
Best Local Similarity	100.0%	Pred. No. 5.6e-185		
Matches 219	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MKLLWACTIVCAVAFARRRRFPGEDNDGHPHLSLNTPYGIRNLPPLYYRPNVTVP	60	
DB	1	MKLLWACTIVCAVAFARRRRFPGEDNDGHPHLSLNTPYGIRNLPPLYYRPNVTVP	60	
QY	61	SYRNTYTDGTLGSPYILTSPPGPPYHNRGPPPLAQLVNPPPPGPFVPSRRFSA	120	
DB	61	SYRNTYTDGTLGSPYILTSPPGPPYHNRGPPPLAQLVNPPPPGPFVPSRRFSA	120	
QY	121	AAAPAPPIAEPAAAPLTAATPAAPPAAGAPVPAAPPAAPPAAPPAAPPAAPPA	180	
DB	121	AAAPAPPIAEPAAAPLTAATPAAPPAAGAPVPAAPPAAPPAAPPAAPPAAPPA	180	
QY	181	EAPVGPAAEPSPAPAPATAKPAAPHPSPLEQANQ	219	

Db 181 EAPVGVEPAAEPSPAEPATAKPAAPERHPSPSLQANQ 219

```

RESULT 2
US-09-923-236-2
: Sequence 2, Application US/09923236
: Patent No. US20020090677A1
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Adler, David A.
: TITLE OF INVENTION: SECRETED SALIVARY ZS1663 POLYPEPTIDE
: FILE REFERENCE: 97-71
: CURRENT APPLICATION NUMBER: US/09/923.236
: CURRENT FILING DATE: 2001-08-03
: PRIOR APPLICATION NUMBER: US 60/124,820
: PRIOR FILING DATE: 1999-03-17
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FASTSEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 219
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-923-236-2

```

Query March	100.0%;	Score 219;	DB 9;	Length 219;
Best Local Similarity	100.0%;	Pred. No. 5.6e-185;		
Matches 219;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

Qy	1	MKLLIACIVCAFAFKRRFPFGEEDNDGGHLPSPSNTIPYGRINLPPPLYTPVNTVP	60
Db	1	MKLLIACIVCAFAFKRRFPFGEEDNDGGHLPSPSNTIPYGRINLPPPLYTPVNTVP	60
Qy	61	SYDGNITYTDIGLPSYPMILTSBPFYYHIRGFLPLATQNLNPPLEPRGPFPVPSRFFSA	120
Db	61	SYDGNITYTDIGLPSYPMILTSBPFYYHIRGFLPLATQNLNPPLEPRGPFPVPSRFFSA	120
Qy	121	AAAPAAPPIAEEPAALPTATPVAAEPAGAPVAEPPAAEPVGAEPAAEAIPVAAEPAA	180
Db	121	AAAPAAPPIAEEPAALPTATPVAAEPAGAPVAEPPAAEPVGAEPAAEAIPVAAEPAA	180
Qy	181	EAPVGVEPAAEESPAPETAKPAAPEPHPSLSLEQAQQ	219
Db	181	EAPVGVEPAAEESPAPETAKPAAPEPHPSLSLEQAQQ	219

```

RESULT 3
US-09-922-469-2
: Sequence 2, Application US/093922469
: Patent No. 'US20020173072A1
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.
: APPLICANT: Adler, David A.
: TITLE OF INVENTION: SECRETED SALIVARY ZSIC63 POLYPEPTIDE
: FILE REFERENCE: 97-71
: CURRENT APPLICATION NUMBER: US/09/922,469
: CURRENT FILING DATE: 2001-08-03
: PRIOR APPLICATION NUMBER: US 60/124,820
: PRIOR FILING DATE: 1999-03-17
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: PASTESEQ for Windows Version 3.0
: SEQ ID NO 2
: LENGTH: 219
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-922-469-2

```

```

Query Match      100.0%; Score 219; DB 10; Length 219;
Best Local Similarity 100.0%; Pred. No. 5.6e-185;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY      1 MKLLIMATIVCAAPARKRRFPPIGSDNDGDGRLPRLPSLNIPIGIRNLPPPLYYRPNVYP 60
|||||

```

Db	1	MKLLIMACIVCAFAFKRRKRPPIGSDDDNDGHRPLHPSLNI PYGIRNLPPRLYYRPVNTVB	60
Qy	61	SYPGNTYTDGLPSYPMILITSGBPYYVYHNGFPLATQIANVPLPPRGPFYLPSPSFFGA	120
Db	61	SYPGNTYTDGLPSYPMILITSGBPYYVYHNGFPLATQIANVPLPPRGPFYLPSPSFFGA	120
Qy	121	AAAPAAPPIAEBPAAABLTATPVAAEBPAAGAPVAABEPAAEA PVABEPAA	180
Db	121	AAAPAAPPIAEBPAAABLTATPVAAEBPAAGAPVAABEPAAEA PVABEPAA	180
Qy	181	EAPVGVBPAAEBSPABPATAKPAAPBPHSPSLQDANQ	219
Db	181	EAPVGVBPAAEBSPABPATAKPAAPBPHSPSLQDANQ	219

```

RESULT 4
US-10-231-417-192
; Sequence 192, Application US/10231417
; Publication No. US20030176681A1
GENERAL INFORMATION:
; APPLICANT: Feng et al.
; TITLE OF INVENTION: 148 Human Secreted Proteins
; FILE REFERENCE: P2019p1
CURRENT APPLICATION NUMBER: US/10/231,417
CURRENT FILING DATE: 2002-08-30
PRIOR APPLICATION NUMBER: US/09/296,622
PRIOR FILING DATE: 1999-04-23
NUMBER OF SEQ ID NOS: 619
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 192
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
FEATURE:
; NAME/KEY: SITE
; LOCATION: (1159)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
FEATURE:
; NAME/KEY: SITE
; LOCATION: (221)
OTHER INFORMATION: Xaa equals stop translation
US-10-231-417-192

```

Query Match	68.5%	Score 150;	DB 12;	length 221;
Best Local Similarity	100.0%	Pred. No. 3.3e-124;		
Matches 150;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	MKLLIMACTVCVAFARAKRRPPIGSDNDNDGHPRLSPINIPGIRNLPEPLLYRPNVNTVP	60
Db	1	MKLLIMACTVCVAFARAKRRPPIGSDNDNDGHPRLSPINIPGIRNLPEPLLYRPNVNTVP	60
Qy	61	SYPGNTYDNDLSAPYPMILTSBGPYYVYHNGFPLATOLNVPLPRGSPFYPSKFFSA	120
Db	61	SYPGNTYDNDLSAPYPMILTSBGPYYVYHNGFPLATOLNVPLPRGSPFYPSKFFSA	120
Qy	121	AAAPAPPIAAEPAAAPLTATPVAAEPAA	150
Db	121	AAAPAPPIAAEPAAAPLTATPVAAEPAA	150

RESULT 5
US-09-746-783-106
Sequence 106, Application US/09746783
Publication No. US20030044935A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Racle, Lisa A.
Trecay, Maurice
Spaulding, Vikki
Agostino, Michael J.
Howes, Steven H.

```

; Fecheel, Kim
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; ENCODING THEM
; NUMBER OF SEQUENCES: 231
; CORRESPONDENCE ADDRESSES:
; ADDRESSER: Genetics Institute, Inc.
; STREET: 87 Cambridge Park Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/746,783
; FILING DATE: 21-Dec-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Mlaasincic, Debra J.
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 219 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 106:
US-09-746-783-106

Query Match 63.9%; Score 140; DB 11; Length 219;
Best Local Similarity 100.0%; Pred. No. 2.1e-115;
Matches 140; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKLIMACTIVVAPARKRRFPFEGDDNDGHPHPSLNTYGRNLPPPLYTPPVNTVP 60
DB 1 MKLIMACTIVVAPARKRRFPFEGDDNDGHPHPSLNTYGRNLPPPLYTPPVNTVP 60
QY 61 SYPGNTYDTGLPSYPMILTSRGPYVYHIRGPIATQNLNPPPLPPRGFPVPSRPFSA 120
DB 61 SYPGNTYDTGLPSYPMILTSRGPYVYHIRGPIATQNLNPPPLPPRGFPVPSRPFSA 120
QY 121 AAAPAPPIAAEPAAAPLT 140
DB 121 AAAPAPPIAAEPAAAPLT 140

RESULT 6
US-10-156-761-13162
; Sequence 13162, Application US/10156761
; Publication No. US20030113018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13162
; LENGTH: 798
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13162

Query Match 4.1%; Score 9; DB 15; Length 798;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 198 PATAKPAAP 206
DB 29 PATAKPAAP 37

RESULT 7
US-09-738-626-4434
; Sequence 4434, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIJO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4434
; LENGTH: 2969
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4434

Query Match 4.1%; Score 9; DB 10; Length 2969;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 PAAAPAPVA 176
DB 1663 PAAAPAPVA 1671

RESULT 8
US-10-289-762-96
; Sequence 96, Application US/10289762
; Publication No. US20040006218A1
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragment,
; and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/10/289,762
; PRIOR FILING DATE: 2003-03-27
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 96
; LENGTH: 132

```

TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-10-289-762-96

Query Match 3.7%; Score 8; DB 12; Length 132;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAEP 178
DB 53 EAPVAAEP 60

RESULT 9
US-10-219-220-250
Sequence 250, Application US/10219220
Publication No. US20030082724A1
GENERAL INFORMATION:
APPLICANT: Lasham, Barry
APPLICANT: Lasham, Annette
TITLE OF INVENTION: Compositions affecting programmed cell
TITLE OF INVENTION: death and their use in the modification of plant development
FILE REFERENCE: 11000.102261
CURRENT APPLICATION NUMBER: US/10/219,220
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
PRIOR FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 250
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 250
LENGTH: 206
TYPE: PRT
ORGANISM: Eucalyptus grandis
US-10-219-220-250

Query Match 3.7%; Score 8; DB 15; Length 206;
Best Local Similarity 100.0%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAAP 127
DB 112 AAAAPAAP 119

RESULT 10
US-10-180-375-82
Sequence 82, Application US/10180375
Publication No. US2003012638A1
GENERAL INFORMATION:
APPLICANT: Cahoon, William B.
APPLICANT: Ramodu, Omolayo O.
APPLICANT: Harwell, Leslie T.
APPLICANT: Helentjaris, Timothy
APPLICANT: Li, Changliang
APPLICANT: Lowe, Keith
APPLICANT: Oliveira, Igor Cunha
APPLICANT: Shen, Bo
APPLICANT: Tarczyński, Mitchell C.
TITLE OF INVENTION: Alteration of Oil Traits in Plants
FILE REFERENCE: B01458 US NAI
CURRENT APPLICATION NUMBER: US/10/180,375
CURRENT FILING DATE: 2002-06-26
NUMBER OF SEQ ID NOS: 222
SOFTWARE: Microsoft Office 97
SEQ ID NO 82
LENGTH: 249
TYPE: PRT
ORGANISM: Trifolium aestivum
US-10-180-375-82

Query Match 3.7%; Score 8; DB 15; Length 249;
Best Local Similarity 100.0%; Pred. No. 42;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 VEPAAEP 193
DB 92 VEPAAEP 99

RESULT 11
US-10-369-493-12383
Sequence 12383, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xiandeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR FILING DATE: 2002-02-21
PRIOR APPLICATION NUMBER: US 60/360,039
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 12383
LENGTH: 342
TYPE: PRT
ORGANISM: Aspergillus nidulans
FEATURE:
NAME/KEY: unsure
LOCATION: (1) .. (342)
OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12383

Query Match 3.7%; Score 8; DB 12; Length 342;
Best Local Similarity 100.0%; Pred. No. 55;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 119 SAAAPAA 126
DB 297 SAAAPAA 304

RESULT 12
US-10-369-493-7317
Sequence 7317, Application US/10369493
Publication No. US2003023675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xiandeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 7317
LENGTH: 367
TYPE: PRT
ORGANISM: Burkholderia cepacia
US-10-369-493-7317

Query Match 3.7%; Score 8; DB 12; Length 367;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVAA 156

Db 71 AAGAPVAA 78

```

RESULT 13
US-09-788-345-10
; Sequence 10, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-09-788-345-10

```

Query Match 3.7%; Score 8; DB 10; Length 383;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 354 AAAAPAP 361

```

RESULT 14
US-10-337-312-10
; Sequence 10, Application US/10337312
; Publication No. US20030138451A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/10/337,312
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US/09/788,345
; PRIOR FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 10
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chimeric
US-10-337-312-10

```

Query Match 3.7%; Score 8; DB 12; Length 383;
 Best Local Similarity 100.0%; Pred. No. 60;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 354 AAAAPAP 361

Db 354 AAAAPAP 361

```

RESULT 15
US-09-788-345-12
; Sequence 12, Application US/09788345
; Patent No. US20020147321A1
; GENERAL INFORMATION:
; APPLICANT: ALONSO BEDATE, Carlos
; APPLICANT: REQUENA ROLANIA, Jose M.
; APPLICANT: SOTO ALVAREZ, Manuel
; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
; FILE REFERENCE: BEDATE-1A
; CURRENT APPLICATION NUMBER: US/09/788,345
; CURRENT FILING DATE: 2001-02-21
; PRIOR APPLICATION NUMBER: 09/219,306
; PRIOR FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 12
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal
US-09-788-345-12

```

Query Match 3.7%; Score 8; DB 10; Length 412;
 Best Local Similarity 100.0%; Pred. No. 64;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
 Db 383 AAAAPAP 390

Search completed: January 15, 2004, 15:35:27
 Job time : 33 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 15:28:37 ; Search time 21 Seconds
(without alignments)
441.242 Million cell updates/sec

Title: US-09-923-236-2

Perfect score: 219
Sequence: 1 MKLLIMACIVCAFAFKRRF.....TAKPAAPHPSPSLEQNNQ 219

Scoring table: OLIGO
Gapop 60.0, Gapext 60.0

Searched: 328717 seqs, 42310858 residues

Word size: 0

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	219	100.0	219	US-09-527-345-2	Sequence 2, Appli
2	9	4.1	322	US-08-428-414A-2	Sequence 2, Appli
3	8	3.7	132	US-09-198-452A-96	Sequence 96, Appli
4	8	3.7	168	US-09-252-991A-1819	Sequence 1819, A
5	8	3.7	190	US-09-252-991A-17963	Sequence 17963, A
6	8	3.7	307	US-09-252-991A-21588	Sequence 21588, A
7	8	3.7	383	US-09-471-396-3	Sequence 3, Appli
8	8	3.7	383	US-09-788-345-10	Sequence 10, Appli
9	8	3.7	410	US-09-252-991A-30606	Sequence 30606, A
10	8	3.7	412	US-09-471-396-1	Sequence 1, Appli
11	8	3.7	412	US-09-788-345-12	Sequence 12, Appli
12	8	3.7	549	US-09-252-991A-20196	Sequence 20196, A
13	8	3.7	575	US-08-653-648A-5	Sequence 5, Appli
14	8	3.7	778	US-09-252-991A-27591	Sequence 27591, A
15	8	3.7	4545	US-08-804-227C-14	Sequence 14, Appli
16	8	3.7	4550	US-08-804-227C-8	Sequence 8, Appli
17	8	3.7	4550	US-08-804-198-2	Sequence 2, Appli
18	7	3.2	31	US-09-248-588-27	Sequence 27, Appli
19	7	3.2	76	5273901-11	Patent No. 5273901
20	7	3.2	76	5482709-10	Patent No. 5482709
21	7	3.2	76	5496550-6	Patent No. 5496550
22	7	3.2	79	US-09-252-991A-25964	Sequence 25964, A
23	7	3.2	118	US-08-301-162-10	Sequence 10, Appli
24	7	3.2	118	US-09-461-240-10	Sequence 10, Appli
25	7	3.2	118	US-09-968-927-10	Sequence 10, Appli
26	7	3.2	122	US-09-328-352-7934	Sequence 7934, A
27	7	3.2	134	US-09-252-991A-18886	Sequence 18886, A

28	7	3.2	136	US-09-732-210-814	Sequence 814, App
29	7	3.2	136	US-09-732-210-815	Sequence 815, App
30	7	3.2	141	US-09-252-991A-26122	Sequence 26122, A
31	7	3.2	156	US-08-074-121-5	Sequence 5, Appli
32	7	3.2	156	PCT-US94-06447-5	Sequence 5, Appli
33	7	3.2	159	US-08-481-435-13	Sequence 13, Appli
34	7	3.2	170	US-09-252-991A-27069	Sequence 27069, A
35	7	3.2	173	US-09-252-991A-17373	Sequence 17373, A
36	7	3.2	173	US-09-252-991A-25321	Sequence 25321, A
37	7	3.2	180	5273901-7	Patent No. 5273901
38	7	3.2	180	5482709-6	Patent No. 5482709
39	7	3.2	188	US-09-252-991A-20399	Sequence 20399, A
40	7	3.2	195	US-09-252-991A-20967	Sequence 20967, A
41	7	3.2	197	US-09-252-991A-24218	Sequence 24218, A
42	7	3.2	202	US-09-252-991A-28326	Sequence 28326, A
43	7	3.2	216	US-09-252-991A-18761	Sequence 18761, A
44	7	3.2	239	US-09-252-991A-21250	Sequence 21250, A
45	7	3.2	249	US-09-252-991A-18733	Sequence 18733, A

ALIGNMENTS

```
RESULT 1
US-09-527-345-2
; Sequence 2, Application US/09527345
; Patent No. 6331413
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: SECRETED SALIVARY ZSIG63 POLYPEPTIDE
; FILE REFERENCE: 97-71
; CURRENT APPLICATION NUMBER: US/09/527,345
; PRIOR APPLICATION NUMBER: US 60/124,820
; PRIOR FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 219
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-527-345-2

Query Match      100.0%; Score 219; DB 4; Length 219;
Best Local Similarity 100.0%; Pred. No. 1.1e-203; Indels 0; Gaps 0;
Matches 219; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1 MKLLIMACIVCAFAFKRRFPFEGEDNDGHPHPSLINTPYGIRNLPPPLYRPPVNTVP 60
      |||||
Db      1 MKLLIMACIVCAFAFKRRFPFEGEDNDGHPHPSLINTPYGIRNLPPPLYRPPVNTVP 60

Cy      61 SYPNNTYTDGTPSYPMILTSFGFPYVYHIGPPLATQNLNPPPLPRGFPVPPSRRFSA 120
      |||||
Db      61 SYPNNTYTDGTPSYPMILTSFGFPYVYHIGPPLATQNLNPPPLPRGFPVPPSRRFSA 120

Cy      121 AAAAAPPAAEPAAAPLRTPTVPAABPAAGAPVAAABPAABAPVGAABPAABAAVPAABPA 180
      |||||
Db      121 AAAAAPPAAEPAAAPLRTPTVPAABPAAGAPVAAABPAABAPVGAABPAABAAVPAABPA 180

Cy      181 EAPVGEPPAAEPSPAPATKPAAPHPSPSLEQNNQ 219
      |||||
Db      181 EAPVGEPPAAEPSPAPATKPAAPHPSPSLEQNNQ 219

RESULT 2
US-08-428-414A-2
; Sequence 2, Application US/08428414A
; Patent No. 5912166
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; LEISHMANIASIS
```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/428,414A
FILING DATE: 21-APR-1995
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Kadlecsek, Ann T.
REGISTRATION NUMBER: 39,244
REFERENCE/DOCKET NUMBER: 210121.407
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANDBERRY
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 322 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-414A-2

Query Match 4.1%; Score 9; DB 2; Length 322;
Best Local Similarity 100.0%; Pred. No. 0.73;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 130 AAEPAAAP 138
Db 289 AAEPAAAP 297

RESULT 3
US-09-198-452A-96
Sequence 96, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prevention
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 96
LENGTH: 132
TYPE: PRT
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-96

Query Match 3.7%; Score 8; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 3;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 171 EAPVAAEP 178
Db 53 EAPVAAEP 60

RESULT 4
US-09-252-991A-18199
Sequence 18199, Application US/09252991A

Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18199
LENGTH: 168
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18199

Query Match 3.7%; Score 8; DB 4; Length 168;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 VAAEPAAAG 151.
Db 108 VAAEPAAAG 115

RESULT 5
US-09-252-991A-17963
Sequence 17963, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 17963
LENGTH: 190
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17963

Query Match 3.7%; Score 8; DB 4; Length 190;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAPAAAP 128
Db 113 AAAPAAAP 120

RESULT 6
US-09-252-991A-21588
Sequence 21588, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190

;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 21588
;; LENGTH: 307
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21588

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 307;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 121 AAAAPAP 128
Db 120 AAAAPAP 127

RESULT 7
US-09-471-396-3
;; Sequence 3, Application US/09471396
;; Patent No. 6458359
;; GENERAL INFORMATION:
;; APPLICANT: BEDATE, Carlos Alonso
;; APPLICANT: REQUEENA ROLANTA, Jose Maria
;; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
;; TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
;; TITLE OF INVENTION: L. INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
;; FILE REFERENCE: bedate2a.seq
;; CURRENT APPLICATION NUMBER: US/09/471,396
;; CURRENT FILING DATE: 1999-12-23
;; PRIOR APPLICATION NUMBER: 60/113,825
;; PRIOR FILING DATE: 1998-12-23
;; NUMBER OF SEQ ID NOS: 12
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 3
;; LENGTH: 383
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-3

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 383;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
Db 354 AAAAPAP 361

RESULT 8
US-09-788-345-10
;; Sequence 10, Application US/09788345
;; Patent No. 6525186
;; GENERAL INFORMATION:
;; APPLICANT: ALONSO BEDATE, Carlos
;; APPLICANT: REQUEENA ROLANTA, Jose M.
;; APPLICANT: SOTO ALVAREZ, Manuel
;; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN
;; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG
;; TITLE OF INVENTION: CANINE LEISHMANIOSIS AND PROTEIN OBTAINED
;; FILE REFERENCE: BEDATE-1A
;; CURRENT APPLICATION NUMBER: US/09/788,345
;; CURRENT FILING DATE: 2001-02-21
;; PRIOR APPLICATION NUMBER: 09/219,306
;; PRIOR FILING DATE: 1998-12-23
;; NUMBER OF SEQ ID NOS: 12
;; SOFTWARE: Patentln version 3.1
;; SEQ ID NO 10
;; LENGTH: 383
;; TYPE: PRT

;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: chimeric
US-09-788-345-10

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 383;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
Db 354 AAAAPAP 361

RESULT 9
US-09-252-991A-30606
;; Sequence 30606, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 30606
;; LENGTH: 410
;; TYPE: PRT
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30606

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 410;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 AAGAPVVA 156
Db 37 AAGAPVVA 44

RESULT 10
US-09-471-396-1
;; Sequence 1, Application US/09471396
;; Patent No. 6458359
;; GENERAL INFORMATION:
;; APPLICANT: BEDATE, Carlos Alonso
;; APPLICANT: REQUEENA ROLANTA, Jose Maria
;; APPLICANT: SOTO ALVAREZ, Manuel
;; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE
;; TITLE OF INVENTION: THE ANTIGENIC DETERMINANTS OF FOUR PROTEINS OF
;; TITLE OF INVENTION: L. INFANTUM AND PROTEIN ENCODED BY SAIDI GENE, AND ...
;; FILE REFERENCE: bedate2a.seq
;; CURRENT APPLICATION NUMBER: US/09/471,396
;; CURRENT FILING DATE: 1999-12-23
;; PRIOR APPLICATION NUMBER: 60/113,825
;; PRIOR FILING DATE: 1998-12-23
;; NUMBER OF SEQ ID NOS: 12
;; SOFTWARE: Patentln Ver. 2.1
;; SEQ ID NO 1
;; LENGTH: 412
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Protein Q
US-09-471-396-1

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 412;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 11

US-09-788-345-12
; Sequence 12, Application US/09788345

; Patent No. 6525186

; GENERAL INFORMATION:

; APPLICANT: ALONSO BEDATE, Carlos

; APPLICANT: REQUENA ROLAN, Jose M.

; APPLICANT: SOTO ALVAREZ, Manuel

; TITLE OF INVENTION: CHIMERIC GENE FORMED OF THE DNA SEQUENCES THAT ENCODE THE ANTIGEN

; TITLE OF INVENTION: DETERMINANTS OF FOUR PROTEINS OF L. INFANTUM, USEFUL FOR SEROLOG

; FILE REFERENCE: BEDATE=1A

; CURRENT APPLICATION NUMBER: US/09/788,345

; PRIOR FILING DATE: 2001-02-21

; PRIOR APPLICATION NUMBER: 09/219,306

; PRIOR FILING DATE: 1998-12-23

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patent version 3.1

; SEQ ID NO 12

; LENGTH: 412

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: recombinant sequence from vectors pQ31 and pMal

US-09-788-345-12

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 412;
Pred. No. 8.5;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 120 AAAAPAP 127
|||||

DB 383 AAAAPAP 390

RESULT 12

US-09-252-991A-20196
; Sequence 20196, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 20196

; LENGTH: 549

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20196

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 549;
Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 145 AAEAPAGA 152
|||||

DB 326 AAEAPAGA 333

RESULT 13
US-08-653-648A-5
; Sequence 5, Application US/08653648A

; Patent No. 6379945

; GENERAL INFORMATION:

; APPLICANT: Jepson, Ian

; APPLICANT: Greenland, Andrew

; APPLICANT: Martinez, Alberto

; TITLE OF INVENTION: A Gene Switch

; FILE REFERENCE: PPD50047/US

; CURRENT APPLICATION NUMBER: US/08/653,648A

; PRIOR FILING DATE: 1996-05-24

; PRIOR APPLICATION NUMBER: GB 9510759.5

; PRIOR FILING DATE: 1995-05-26

; PRIOR APPLICATION NUMBER: GB 9605656.9

; PRIOR FILING DATE: 1996-03-18

; PRIOR APPLICATION NUMBER: GB 9513882.2

; PRIOR FILING DATE: 1995-07-07

; PRIOR APPLICATION NUMBER: GB 9517316.7

; PRIOR FILING DATE: 1995-08-24

; NUMBER OF SEQ ID NOS: 65

; SOFTWARE: Patent version 3.0

; SEQ ID NO 5

; LENGTH: 575

; TYPE: PRT

; ORGANISM: Heliothis virescens

US-08-653-648A-5

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 575;
Pred. No. 11;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 TATPVAE 147
|||||

DB 552 TATPVAE 559

RESULT 14

US-09-252-991A-27591
; Sequence 27591, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 27591

; LENGTH: 778

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27591

Query Match
Best Local Similarity 100.0%; Score 8; DB 4; Length 778;
Pred. No. 15;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 148 PAAGPVA 155
|||||

DB 655 PAAGPVA 662

RESULT 15

US-08-804-227C-14
; Sequence 14, Application US/08804227C

; Patent No. 5876991

; GENERAL INFORMATION:

APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhstoss, Stuart A.
APPLICANT: Rostock, Paul R., Jr.
APPLICANT: Sulton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 4545 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-804-227C-14

Query Match 3.7%; Score 8; DB 2; Length 4545;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 131 AEPAAAP 138
DB 2620 AEPAAAP 2627

Search completed: January 15, 2004, 15:34:05
Job time : 21 secs